DIRA : A FRAMEWORK OF DATA INTEGRATION USING DATA QUALITY

Reham I. Abdel Monem¹, Ali H. El-Bastawissy² and Mohamed M. Elwakil³

¹Information Systems Department,

Faculty of computers and information, Cairo University, Giza, Egypt ²Faculty of Computer Science, MSA University, Giza, Egypt ³The Software Engineering Laboratory, Innopolis University, Innopolis, Russia ³Computers and information, Cairo University, Giza, Egypt

ABSTRACT

Data integration is the process of collecting data from different data sources and providing user with unified view of answers that meet his requirements. The quality of query answers can be improved by identifying the quality of data sources according to some quality measures and retrieving data from only significant ones. Query answers that returned from significant data sources can be ranked according to quality requirements that specified in user query and proposed queries types to return only top-k query answers. In this paper, Data integration framework called Data integration to return ranked alternatives (DIRA) will be introduced depending on data quality assessment module that will use data sources quality to choose the significant ones and ranking algorithm to return top-k query answers according to different queries types.

KEYWORDS

Data integration, data sources, query answers, quality measures, top-k query answers, assessment module, ranking algorithm

1. INTRODUCTION

Data integration is the process of combining data from multiple and heterogonous data sources in unified view to satisfy users' queries. It has different architectures but virtual integration and data warehousing architectures are the most commonly used [1]. Data warehouse is a single integrated physical source of data for processing information and it loads data through (ETL) extract, transform and load process, Virtual data integration is the process of combining several local data sources to form single virtual data source. In virtual data integration, data stores in local data sources and accesses through global schema.

Data sources have different levels of quality that specify their fitness for using in specific task, these quality levels change over time and can be measured through some data quality measures that have different classifications [2] and following we will present one of their classifications in table 1.

Data Integration Components	IQ Criteria
Data Source	Reputation,
	Verifiability,
	Availability and
	Response Time.
Schema	Schema
	Completeness,
	Minimalism and
	Type Consistency.
Data	Data
	Completeness,
	Timeliness,
	Accuracy and Data
	Validity.

Table 1.Illustrates data integration IQ measures classification.[3]

Data quality measures can be assessed at different granularities. First, collection of data sources level which assesses the aggregate quality for collection of data sources. Second, data source level which assesses the quality for the whole source. Third, relation level which assesses the quality for a relation attributes and we will use all these levels in our framework. They have relationships between them, these relationships are critical for effective knowledge discovery and finding these relationships or dependencies is dependency discovery. For example, the valid values must be complete values but complete values can be valid values or not.[2]

Following we will focus on data quality measures that could affect the data integration process, could be considered important from user's prospective and we will refer to them as data quality features.

1.1 Data Completeness

Data completeness classified in literature into two types: Null-Completeness and Population Completeness. Null-Completeness is "the degree of missing data or knowing of null values for some data". Population-Completeness is "the availability of all needed data by user" and can be classified into two types of relational model named Closed World Assumption (CWA) and Open World Assumption (OWA). In our work, we will use Population-Completeness under OWA and we will introduce a new type of completeness called Fact-Completeness.

Following, we will introduce the way to measure each type of completeness at attribute level:

Null-Completeness Assessment (C_{Null}): it is the ratio between the number of non-null values (Complete values) and the total number of values or the complement value to ratio between the number of null values (InComplete values) and the total number of values. [4]

$$C_{Null} = \frac{\text{Number of non-null values}}{\text{total number of values}}$$
(1)

Or

$$C_{Null} = 1 - \frac{\text{Number of null values}}{\text{total number of values}}$$
(2)

• Scaled Aggregate Data Completeness Value for Queried Attributes (C)

Scaled Total (C) =
$$\frac{\sum_{1}^{m} C_{Null}(a_m)}{M}$$
 (3)

Where M is total number of queried attributes

• Population-Completeness Assessment ($C_{Population}$): It is the ratio of tuples actually represented in a relation r, with respect to the whole number of tuples in ref(r) where ref(r), is the relation containing all tuples that satisfy the relational schema of r.[4]

$$C_{Population} (\mathbf{r}) = \frac{Cardinality of r}{Cardinality of ref(r)}$$
(4)

1.2 Data Validity

Data validity is "the degree to which attribute value follows specified domain, data item isn't valid if its value is out of the domain" and can be measured at attribute level in our framework.

Data Validity Assessment $(P_{Q_{Validity}})$: It is the ratio between the number of valid values and the total number of values.[4]

$$P_{QValidity} = \frac{\text{Number of valid values}}{\text{total number of values}}$$
(5)

• Scaled Aggregate Data Validity Value for Queried Attributes (L)

Scaled Total (L) =
$$\frac{\sum_{1}^{m} L(a_m)}{M}$$
 (6)

Where M is total number of queried attributes

1.3 Data Accuracy

Data accuracy classified in literature into two types: semantic accuracy and (0 or 1) accuracy. Semantic accuracy refers to the degree of closeness between value v (recorded value) near to value v' (correct value), (0 or 1) accuracy will consider data values are accurate if they don't conflict with real-world values and inaccurate otherwise. In our work, we will use (0 or 1) accuracy and it will be measured at attribute level.

Data Accuracy Assessment $(P_{QAccurate})$: It is the ratio between the number of accurate values and the total number of values.[4]

$$P_{Q_{Accurate}} = \frac{\text{Number of accurate values}}{\text{total number of values}}$$
(7)

• Scaled Aggregate Data Accuracy Value for Queried Attributes (A)

Scaled Total (A) =
$$\frac{\sum_{1}^{m} A(a_m)}{M}$$
 (8)

Where M is total number of queried attributes

1.4 Data Timeliness

Data timeliness is " the degree to which data is up-to-date". So, it captures the gap from data creation to data delivery and can be measured at attribute level in our framework.

• Data Timeliness Assessment: Timeliness is assessed and rescaled according to below equations.[4]

$$Currency = Age + (DeliveryTime - InputTime)$$
(9)

Currency: The degree to which data value reflects all changes that happen to it. Age: How old the data is when it is received. DeliveryTime: The time when data is delivered to user. InputTime: The time when data is obtained.

Timeliness = max
$$\left\{ 0, 1 - \frac{Currency}{Volatility} \right\}$$
 (10)

Volatility: The length of time that data remains valid.

In our work, we will suppose that DeliveryTime = InputTime (no delay from obtaining data to deliver it to user) so Currency = Age

• Aggregate Data Timeliness Value for Queried Attributes (T)

$$Total(T) = \text{Maximum}(T(a_m)) \tag{11}$$

This paper is organized as follows; section 2 will include different approaches concerned with data integration in terms of data quality, the proposed framework for data integration will be explained in section 3. The conclusion and future work will be presented in section 4.

2. RELATED WORK

Many approaches are developed to introduce data integration in terms of data quality. Following, we will present an overview of some approaches related to our framework; how they measure and store data quality, how they process queries and user interference option.

2.1 DaQuinCIS Approach

This approach designed to deal with cooperative information systems and to exchange not only intended data but also metadata. The query processing approach implemented by DaQuinCIS to return a query answer is structured as following:[4]:

- 1. Query unfolding: Each concept in user query Q that sends in terms of global schema is defined in terms of local schemas to retrieve all data that answers user query from all available participating data sources. So, Q will decompose into $Q_1,...,Q_n$ queries to send to each relevant local data source to return results $R_1,...,R_n$.
- 2. Extensional checking: In this step $R_1 \cup R_2 \dots \cup R_n$ are passed to record matching algorithm to discover the same objects. The output of this step is clusters of similar objects.

3. Result building: In this step the best quality object representative will be chosen according to quality value q associated with each field value f. If an object contains the highest quality values for all fields, so it will be chosen as representative otherwise a representative object will be constructed from combination of highest qualified fields' values within cluster. Once all representatives are chosen, the final result will be built from union of all these representatives.

This approach depends on data sources metadata to improve query answers through improving fusion process.

2.2 Data Quality based Data Integration Approach

This approach explains the importance of data quality in data integration. It adds quality system components to integrate data quality dimensions (completeness, accuracy, cost and response time) to data integration system for selecting less number of single data sources for more qualified query results.

This approach presents experiments using Amalgam and THALIA benchmarks to show that the query results delivered in a reasonable amount of time and using the minimum number of possible data sources.[1]

The concept of this approach will be used but to retrieve highest top-k qualified query results from significant data sources only according to different proposed queries types.

3. DATA INTEGRATION TO RETURN RANKED ALTERNATIVES (DIRA) FRAMEWORK

In this section, we will illustrate our new data integration framework called DIRA. DIRA data quality assessment module will be presented in section 3.1, new completeness type (Fact-Completeness) will be explained in section 3.2, DIRA quality system components will be introduced in section 3.3 and DIRA workflow components will be explained in detail in section 3.4.

3.1 DIRA Data Quality Assessment Module

This module consists of different components to evaluate data quality, these components are[5]:

- Assessment Metrics are procedures for calculating data quality features and estimates an assessment score for these features using a scoring function.
- Aggregation Metrics are procedures for calculating aggregated score from distinct assessment scores using aggregation functions like sum, count, and average functions.
- **Data Quality Features** are meta-data for providing user with indication of how data fit to task at hand.

• Scoring Functions are the way for calculating data quality features. There are many scoring functions to choose between them like simple comparisons, set function, aggregation function and complex statistical function.

3.2 Data Fact-Completeness

Data fact-completeness is a new accurate type of completeness where it uses null-completeness and population-completeness to assess its value.

• Fact-Completeness Assessment (C_{fact}) : It is subtraction of the probability of incomplete values from the probability of population-completeness values. We will present its equation on attribute level and then we will aggregate the values for higher levels.

$$C_{Population}(a_m(\mathbf{r})) = \frac{Cardinality of a_m(\mathbf{r})}{Cardinality of ref(r)}$$
(12)

$$InC_{Null}(a_m(\mathbf{r})) = \frac{Count of null values for a_m}{Cardinality of ref(r)}$$
(13)

$$C_{fact}(a_m(\mathbf{r})) = C_{Population}(a_m(\mathbf{r})) - InC_{Null}(a_m(\mathbf{r}))$$
(14)

Where a_m (r) refers to attribute number m in relation r

• Scaled Aggregate Data Fact-Completeness Value for Queried Attributes (C)

Scaled Total (C) =
$$\frac{\sum_{1}^{m} c_{fact}(a_m)}{M}$$
 (15)

Where M is total number of queried attributes

3.3 DIRA Quality System Components

In our work, we will add some components to integration systems called quality system components to improve query answers. These components are data quality acquisition and user input.

3.3.1 Data Quality Acquisition

This component is responsible for storing attributes and relations found in data sources in metadata store. It is also responsible for running data quality queries and storing their answers in the metadata store

DIRA Metadata Store that is presented in figure1 will use the concept of hierarchical quality framework [6] to build its entities that we will explain in table 2.



International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.6, No.2, March 2016

Figure 1.DIRA metadata structure

Table 2.Presents DIRA	🛾 metadata	structure	entities
-----------------------	------------	-----------	----------

Entity	Definition			
DataSource	This entity contains information about all data sources			
	participating in data integration process.			
Domain	This entity contains information about data sources' domains.			
Table	This entity contains information about all data sources' tables.			
Column	This entity contains information about all tables' columns.			
GlobalSchemaTable	This entity contains information about all tables in global			
	schema.			
GlobalSchemaColumn	This entity contains information about all columns in every			
	table in global schema.			

GlobalSchemaMapping	 This entity is associative entity It contains information about tables' columns with their correspondence in global schema. It contains scores that evaluate the data quality for every data source column with its correspondence in global schema according to scoped data quality features, these scores will be calculated once during data integration system configuration, they will be updated according to data sources modification and they will be used in evaluating the results that will return from queried data sources (data integration will retrieve data from only sources that can answer query and can satisfy the required level of quality). Scores assessment will save time and cost for data integration process especially for data sources with high volatility.
QueriedDataSource	This entity contains information about data sources that can participate with attributes in query answering and which attributes it can participate with.
QueriedDataSourceAssessmentMetric	This entity contains information about every data source that can participate in query answering and its data quality features' total scores (this entity represents evaluation for data that every data source can participate with in answering query).
AlternativeAggregatedMetric	 This entity contains information about qualified alternatives (qualified alternative is one or more queried data source that can answer query and can satisfy the required level of quality if specified) for given query. It contains qualified alternatives data quality features scores (aggregated scores for alternatives with more than one data source and assessment scores for alternatives with one data source). Qualified alternatives will pass to ranking algorithm to return top-k ranking alternatives before duplicate detection and data fusion.
QueriedDataSourceAssessmentMetric_ AlternativeAggregatedMetric	This entity is associative entity that contains the IDs of qualified alternatives aggregated metrics and IDs of their related queried data sources assessment metrics.

3.3.2 User Input

SQL can be extended to include some quality constrains that will be required by user in query to return qualified results, these constrains are expressed by data quality features. Query Q syntax with quality constraint [1]

Select A1 Ak From G	
Where < selection condition >	
With < data quality goal > Where A1.A2, Ai are global attributes of G	

3.4 DIRA Workflow Components

In this section, we will explain in detail the DIRA workflow components (Data Sources Attributes (columns) Assessment Metrics, Queried Data Sources Assessment Metrics, Alternative Formation, Alternatives Aggregated Metrics and Alternatives Ranking) and they will be presented in figure2.

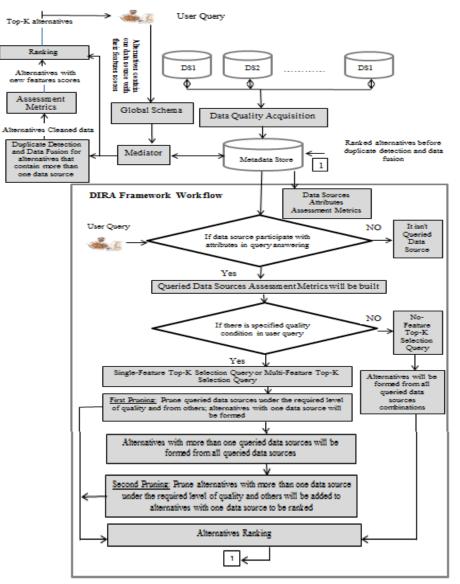
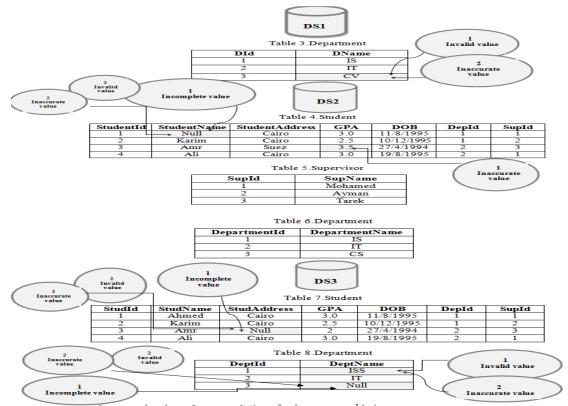
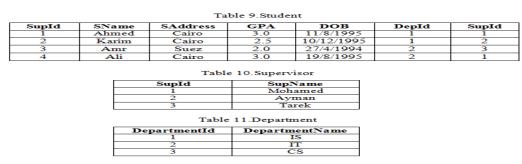


Figure 2.DIRA workflow components

DIRA components will be explained in the following motivation example1. Example1 represents three data sources DS1, DS2 and DS3 with their data and the status of these data from data completeness, data validity and data accuracy (Note: The assessment date was on 2/2/2016). In relation data; one refers to value status and two refers to the consequences of this status.



Suppose that the reference relations for data sources relations are



3.4.1 Data Sources Attributes Assessment Metrics

In this component, we will assess the scoped data quality features scores for all attributes (columns) in data sources with its correspondence in global schema as we presented in global schema mapping entity. Following are tables that we will use to build this component

Table 12.Domain entity

Don

Table 13.DataSource entity

DomainID

1.772		DataSourceId	DataSourceName	
nainID	DomainName	1	DS1	Г
1	Cairo University	2	D52	F
*	cure on enaly	3	DS3	

Table 14. Table entity

TableId	TableName	InsertionDate	Volatility	DataSourceId
1	Department	2/12/2015	365	1
2	Student	2/1/2016	365	2
3	Supervisor	2/1/2016	365	2
4	Department	2/1/2016	365	2
5	Student	2/10/2015	365	3
6	Department	2/10/2015	365	3

Table 16.GlobalSchemaTable entity

GSTableId	GSTableName
1	Student
2	Supervisor
3	Department

Table 15.Column entity

ColumnId	ColumnName	TableId
1	DId	1
2	DName	1
3	StudentId	2
-4	StudentName	2
5	StudentAddress	2
6	GPA	2
7	DOB	2
8	SupId	3
9	SupName	3
10	DepartmentId	-4
11	DepartmentName	-4
12	StudId	5
13	StudName	5
1.4	StudAddress	5
15	GPA	5
16	DOB	5
17	DeptId	6
16	DeptName	6

Table 17.GlobalSchemaColumn entity

GSColumnId	GSColumnName	Detector	CorelatedWith	GSTableId
1	SId	0	Null	1
2	SName	1	Null	1
3	SAddress	1	Noll	1
4	GPA	0	Null	1
5	DOB	1	Null	1
6	SupId	0	Null	2
7	SupName	0	Null	2
8	DId	0	Null	3
9	DName	0	Null	3

According to equations 5, 7, 9, 10, 12, 13 and 14 that we presented in section 1 and section 3.2, the GlobalSchemaMapping entity will be presented in table 18.

N.P: According to data quality features' relationships[2], we can deduce the following relation (1) to validate values in table 18.

$$P_{Q_{Completeness}} \ge P_{Q_{Validity}} \ge P_{Q_{Accuracy}} \tag{1}$$

Column	GSColumn	Column	Column	Column	Column	Column	Column
Id	Id	Population	InCompletness	FactCompletness	Validity	Accuracy	Timeliness
		Completness					
3	1	1	0	1	1	1	0.92
12	1	1	0	1	1	1	0.67
4	2	1	0.25	0.75	0.75	0.75	0.92
13	2	1	0	1	1	1	0.67
5	3	1	0	1	1	1	0.92
14	3	1	0.25	0.75	0.75	0.75	0.67
6	4	1	0	1	1	0.75	0.92
15	4	1	0	1	1	1	0.67
7	5	1	0	1	1	1	0.92
16	5	1	0	1	1	1	0.67
8	6	1	0	1	1	1	0.92
9	7	1	0	1	1	1	0.92
1	8	1	0	1	1	1	0.84
10	8	1	0	1	1	1	0.92
17	8	1	0	1	1	1	0.67
2	9	1	0	1	0.67	0.67	0.84
11	9	1	0	1	1	1	0.92
18	9	1	0.33	0.67	0.33	0.33	0.67

Table 18. Global Schema Mapping entity

3.4.2 Queried Data Sources Assessment Metrics

In this component first, we will fill QueriedDataSource associative entity table 19 with data related to each query.

	Table 19.QueriedDa	ataSource		
	QueriedDataNourceld	DataSourceld	ColumnId	GacolumnId
	1	2	-1	2
	2	3	1.3	2
Q1: Select SName, SAddress, DOB, SupName, DName	3	2	5	3
Ar and a sumple and a sol a shown a sub-	1	э	1.4	3
From G	5	2	7	5
roud	0	3	1.0	
	7	2	9	7
			2	9
	9	2	11	9
	10	3	18	9

Second, we will fill QueriedDataSourceAssessmentMetric entity table 20 with data.

International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.6, No.2, March 2016

Queried	Queried	Queried	Queried	Queried	Queried	Queried	Queried
DataSource	Column	GSColumn	GSColumn	Column	Column	Column	Column
Id	Id	Id	Name	FactCompleteness	Validity	Accuracy	Timeliness
1	2	9	DName	1	0.67	0.67	0.84
2	4	2	SName	0.75	0.75	0.75	0.92
2	5	3	SAddress	1	1	1	0.92
2	7	5	DOB	1	1	1	0.92
2	9	7	SupName	1	1	1	0.92
2	11	9	DName	1	1	1	0.92
3	13	2	SName	1	1	1	0.67
3	14	3	SAddress	0.75	0.75	0.75	0.67
3	16	5	DOB	1	1	1	0.67
3	18	9	DName	0.67	0.33	0.33	0.67

Table 20.Illustrative table

According to equations 6, 8, 11 and 15 that we presented in section 1 and section 3.2, the QueriedDataSourceAssessmentMetric entity for Q1 will be presented in table 21

QueriedDS Assessment MetricId	Queried Data Source Id	Asses Metric	idDS sment Name	QueriedDS Assessment Metric Definition	QueriedDS Assessment Metric Temporary	QueriedDS Assessment MetricState	QueriedDS Assessment Metric History
1	1	Metric		This metric is evaluation for DS1 required attributes data quality	Dynamic	Positive	This evaluation metric is for DS1 that found in Queried DataSource entity with its required attributes
2	2	DS2Assessment Metric		This metric is evaluation for DS2 required attributes data quality	Dynamic	Positive	This evaluation metric is for DS2 that found in Queried DataSource entity with its required attributes
3	3	DS3Assessment Metric		This metric is evaluation for DS3 required attributes data quality	Dynamic	Positive	This evaluation metric is for DS3 that found in Queried DataSource entity with its required attributes.
		QueriedDS Assessment Metric Id	Information Supplier	QueriedDS Fact Completeness	QueriedDS Validity	QueriedDS Accuracy	QueriedDS Timeliness
		1	User	0.20	0.13	0.13	0.84
		2	User	0.95	0.95	0.95	0.92
		3	User	0.68	0.62	0.62	0.67

Table 21.QueriedDataSourceAssessmentMetric entity

3.4.3 Alternatives Formation

Users became not interested in how to access different data sources or how to combine the results from them. Users' requested data can be found in single source, in different sources or distributed between many sources.

<u>Alternative</u> represents one or more queried data source that participates in data integration, it may be qualified alternative or not qualified alternative. <u>N/P:</u> In our framework,

• Alternatives formation will be specified according to query type.

- For queries without any required quality condition, alternatives will be formed from combinations of all queried data sources and all will be considered as qualified alternatives.
- For queries with quality condition, we will consider every queried data source that satisfies the required quality level as qualified alternative of single queried data source and we will prune others from forming alternatives from one queried data source (First Pruning).We will build combinations from all queried data sources to form alternatives from two or three or more queried data sources, alternatives that will not satisfy the required level of quality will be pruned (Second Pruning) and the remains will be considered as qualified alternatives.
- Total number of alternatives before first and second pruning will be within $\{0, \dots, 2^M 1\}$.

Alternatives formation according to Q2 (Q2 with quality condition)



First pruning according to Q2: from QueriedDataSourceAssessmentMetric table 21, we can specify that DS1 will prune from forming alternative alone (not qualified alternative) as presented in table 23 because it is under the level of quality specified in user query Q2.

Qualified alternatives with one queried data source: Alternative2: (DS2), Alternative3: (DS3)

Table 22.QueriedDataSourceAssessmentMetric_AlternativeAggregatedMetric entity for Q2 after first pruning

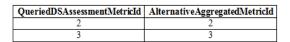


Table 23.AlternativeAg		C 00 C C	
I able 73 Alternative Ad	areasted Metric entit	v tor () ottor ti	rot nruning

Alternative Aggregated MetricId	Alternative Aggregated MetricName	Alternative Aggregated MetricDefinition	Alternative Aggregated MetricTemporary	Alternative Aggregated MetricState	Aggr Mo His	native egated etric story	Information Supplier
2	Alternative2	This alternative contains DS2Assessment Metric and it can satisfy the required level of data quality where query answer completeness≥0.65	Dynamic	Positive	This alterns evalua Ds2 Assess Metric found Querie Source Assess Metric	ited by ment that in dData ment	User
3	Alternative3	This alternative contains DS3Assessment Metric and it can satisfy the required level of data quality where query answer completeness≥0.65	Dynamic	Positive	This alterns evalua Ds3 Assess Metric Source Assess Metric	ited by ment that in dData ment	User
Alternative Aggregated MetricId	Alternative Fact Completeness	Alternative Validity	Alternative Accuracy	Alternativ Timeline			
2	0.95	0.95	0.95	0.92			
3	0.68	0.62	0.62	0.67			

Second pruning according to Q2: from QueriedDataSourceAssessmentMetric table 21, we will build combinations from queried data sources to form alternatives from two or three or more queried data sources as we will introduce in table 24 and we will prune alternatives that will not satisfy the required level of quality according to their calculated aggregated metrics that will presented in section 3.4.4.

Alternatives from two or more queried data sources: Alternative4: (DS1, DS2), Alternative5: (DS1, DS3), Alternative6: (DS2, DS3) and Alternative7: (DS1, DS2, DS3)

Table 24.QueriedDataSourceAssessmentMetric_AlternativeAggregatedMetric entity for Q2 before second punning

QueriedDSAssessmentMetricId	AlternativeAggregatedMetricId
2	2
3	3
1	4
2	4
1	5
3	5
2	6
3	6
1	7
2	7
3	7

3.4.4 Alternatives Aggregated Metrics

In this component, aggregated scores will be calculated for alternatives with two or more queried data sources from their assessment metrics.

Following, we will present equations to assess our scoped data quality features aggregated scores for alternatives where Q is number of data sources that form the alternative[7]

• The fact completeness of alternative (collections of DBs)

$$C_{DBS} = \sum_{q=1}^{Q} C_{fact}(DS_q) / Q \tag{16}$$

• The data validity of alternative (collections of DBs)

$$L_{DBS} = \sum_{q=1}^{Q} L(DS_q) / Q \tag{17}$$

• The data accuracy of alternative (collections of DBs)

$$A_{DBS} = \sum_{q=1}^{Q} A(DS_q) / Q \tag{18}$$

• The data timeliness of alternative (collections of DBs)

$$TotalAlternativeT = Maximum(T(DS_a)) in Alternartive)$$
(19)

By applying equations 16, 17, 18 and 19 on Queried Data Sources Assessment Metrics table 21 according to Q2, The Alternatives Aggregated Metrics for Alternative4, Alternative5, Alternative6 and Alternative7 will be presented in table 25

Alternative Meterical M		Alternative		Alternative			
Mericidal Material Schement Metric Temperary Meint Status Metric	Alternative		Alternative		Alternative	Alternative	Information
2 Alternative2 This alternative exclusted DA between the second of the				MetricTemporary	MetricState		Supplier
ensisteness Delaware des quary ensisteness des quary (arry) ensisteness des quary (arry) ensisteness des quary (arry) ensisteness des quary (arry) ensisteness des quary (arry) ensisteness des quary (arry) Dynamic Perside This des mainterest des quary User (arry) 5 Atternatived This enclasse des mainterest des mainterest des quary Dynamic Perside This destinaterest des quary User 4 Atternatived This enclasse des quary Dynamic Perside This destinaterest des quary User 4 Atternatived This enclasse des quary Dynamic Perside This destinaterest des quary User 4 Atternatived This enclass D13 des quary Dynamic Perside This des quary User 3 Atternatived This ensites D13 des quary Dynamic Perside This des quary User 5 Atternatived This ensites D21 des quary Dynamic Perside This des quary User 6 Atternatived This diamative des quary Dynamic Perside This des quary User				Dynamic			User
Alternatived This alternative Dynamic Particle Meric shafe found in Control of decimation of the state of			contains				
de. required twol, of constant quary, where quary, and of the required service quark, where quark and the required service quark, where quark, where quark and the required service quark and the required service quark, where quark and the required service quark and the requi					1		
data quality market every Quality Quality Summarket Massie entity 5 Alternative2 This Bornative entities Dynamic Fastive This 6 Alternative2 This alternative entities Dynamic Pastive This 4 Alternative3 This alternative entities Dynamic Pastive This 4 Alternative3 alternative entities Dynamic Pastive This Blemative dueUser 4 Alternative3 alternative entities Dynamic Pastive This Blemative dueUser 4 Alternative3 alternative entities Dynamic Pastive This Blemative dueUser 5 Alternative3 This alternative entities Dynamic Pastive This Blemative dueUser 5 Alternative3 This alternative dueDynamic Pastive This Blemative dueUser 6 Alternative3 This alternative dueDynamic Pastive This Blemative dueUser 7 Alternative3 This alternative dueDynamic Dynamic PastiveThis User 7 Alternative3 This					1	Metric that found	
query assure Source Source 5 Alternative3 This alternative Dynamic Positive relative			the required level of		1		
S Alternatived This alternative D33 Assessment Metric and it can satisfy draw regulated with of a draw regulated w					1		
3 Alternatived This alternative entities alternative entalternative entalternative entities alternative entities					1		
2 Alternative2 This alternative centains Dynamic Fusion This alternative austed by Dot Assessment States and Constitution User 4 Alternatived This alternative data Dynamic Positive This alternative data User 4 Alternatived This alternative data Dynamic Positive This alternative data User 4 Alternatived This alternative data Dynamic Positive data This alternative data User 4 Alternatived This alternative data Dynamic Positive data This alternative data User 5 Alternatived This alternative data Dynamic Positive data This alternative data User 6 Alternatived This alternative data Dynamic Positive This alternative data User 7 Alternative? This alternative data Dynamic Positive This alternative data User 6 Alternative? This alternative data Dynamic Positive This alternative data User 7 Alternative? This					1		
4 Alternatived This alternative error tails and it can study dam quality where quary with where quary quality where quary and an error tails of the second state of the second quary and the second q		Alternative3	This alternative	Dynamic	Penitive		Tree
Alternative Metric and it can satisfy the required level of data quality where years, a guality where years, a model Data a second second complements 200 second complements 200 second a second contains D51 and D52 Assessment Metrics Metric that found a model Data Second Second a second contains D51 and D52 Assessment Metrics 4 Alternatives ⁴ This a lisenative contains D51 and D52 Assessment Metrics Dynamic Pesitive a second contains the found in QueriedData Second Second D53 Assessment Metrics User 3 Alternatives ⁴ This a lisenative contains D51 and D53 Assessment Metrics Dynamic Pesitive Pesitive Assessment Metric and by D51 and D53 Assessment Metrics User 6 Alternative ⁶ This a lisenative contains D52 and D53 Assessment Metrics Dynamic Pesitive Pesitive This alternative Assessment Metric antip to positive assessment Metric antip to positive antip to contains D51, D52 and D53 Assessment Metrics Dynamic Pesitive This alternative assessment Metric antip to positive assessment Metric antip to positive assessment Metric antip to positive antip to contains D51, D52 and D53 Assessment Metrics User 7 Alternative ⁷ This alternative Completences Dynamic Pesitive to positive to positi			contains			evaluated by Da3	
des required level of data quairver complianenze of data quairver complianenze 4 in GuarialDus Seuser Anternative en be evaluated by DS1 and DS2 Assessment Metric shart base Seuser Assessment Metric shart Seuser Assessment Metric Seuse Assessment Metric Seuse Asses					1		
data query query completenency20.63 Dynamic Pessive server Assessment Motio entity by DS1 and DS2 Assessment Metrics User 4 Alternative4 This entities DS1 and DS2 Assessment Metrics Dynamic Pessive manual set evaluated by DS1 and DS2 Assessment Metrics hat found in QueriedData Saree Assessment Metrics User 5 Alternative5 This entities Dynamic Pessive manual set evaluated by DS1 and DS2 Assessment Metrics User 6 Alternative5 This entities Dynamic Pessive manual set evaluated by DS1 and DS3 Assessment Metrics User 6 Alternative5 This entities Dynamic Pessive manual set evaluated by DS1 and DS3 Assessment Metrics User 7 Alternative6 This elemative contains Dynamic Pessive manual set evaluated bis curre assessment Metrics User 7 Alternative6 This elemative contains Dynamic Pessive manual set evaluated bis curre assessment Metrics User 7 Alternative restains This elemative DS3 Assessment Metrics Dynamic Pessive manual set evaluated by DS1, DS2 and DS3 Assessment Metrics User 7 Alternative restains Alternative Completions Alternative Completions This elemative restains Diser 7 Alternative Alternative Alternative Completions Alternative Completions					1		
query answer Severe 4 Alternative4 This alternative Dynamic Pesitive This alternative 4 Alternative4 This alternative Dynamic Pesitive This alternative 4 Alternative3 This alternative Dynamic Pesitive This alternative 5 Alternative3 This alternative Dynamic Pesitive This blantained 5 Alternative3 This alternative Dynamic Pesitive This blantained 6 Alternative3 This alternative Dynamic Pesitive This blantained 6 Alternative3 This alternative Dynamic Pesitive This blantained 7 Alternative3 This alternative Dynamic Pesitive This blantained 7 Alternative7 This alternative Dynamic Pesitive This <					1		
Alternative Assessment Metric entity Assessment Metric entity Assessment Metric entity User entities 4 Alternative4 This alternative entities Dynamic Pestitive This alternative entity User entity User 3 Alternative3 This alternative entities Dynamic Pestitive This alternative entities User 3 Alternative3 This alternative entities Dynamic Pestitive This alternative entities User 4 Alternative3 This alternative entities Dynamic Pestitive This alternative entities User 5 Alternative3 This alternative centaties Dynamic Pestitive This alternative entities User 6 Alternative6 This alternative centaties Dynamic Pestitive This alternative entities User 7 Alternative7 This alternative centaties Dynamic Pestitive This alternative entities User 7 Alternative7 This alternative centaties Dynamic Pestitive This alternative entisentis Us					1		
4 Alternative4 This alternative erations DS1 and DS2 Assessment Metrics Dynamic Pessive Metric entity 5 Alternative5 This alternative contains DS1 and DS2 Assessment Metrics Dynamic Pessive This alternative alternative for this bill and DS2 Assessment Metrics Dynamic Pessive This alternative for this bill and DS3 Assessment Metrics User 5 Alternative5 This alternative contains DS1 and DS3 Assessment Metrics Dynamic Pessive This alternative contains DS1 and DS3 Assessment Metrics User 6 Alternative6 This alternative contains DS2 and DS3 Assessment Metrics Dynamic Pessive This alternative contains BS2 and DS3 Assessment Metrics Metrics that found in QuericBDa3 Assessment Metrics User 7 Alternative7 This alternative dot DS3 Assessment Metrics Dynamic Pessive This alternative contains BS2 and DS3 Assessment Metrics Dynamic Pessive User 7 Alternative7 This alternative Alte					1		
4 Alternative4 This alternative constained DS1 and DS2 Assessment Metrics Dynamic Peakive This alternative to vulnated by DS1 and DS2 Assessment Metrics User 3 Alternative3 This alternative Dynamic Peakive This alternative 3 Alternative3 This alternative Dynamic Peakive This alternative 4 Alternative3 This alternative Dynamic Peakive This alternative 5 Alternative3 This alternative Dynamic Peakive This alternative 6 Alternative6 This alternative Dynamic Peakive This alternative 6 Alternative6 This alternative Dynamic Peakive This alternative 7 Alternative7 This alternative <td< td=""><td></td><td>1</td><td></td><td></td><td> </td><td></td><td></td></td<>		1					
een tains DS1 and DS2 Assessment Metrics openation ear be evaluated by DS1 and DS2 Assessment Metrics 3 Alternative3 This alternative een tains DS1 and DS3 Assessment Metrics Dynamic Feasible Peasible This alternative end tains DS1 and DS3 Assessment Metrics User 6 Alternative3 This alternative een tains DS1 and DS3 Assessment Metrics Dynamic Peasible This alternative matrix User 6 Alternative3 This alternative een tains DS2 and DS3 Assessment Metrics Dynamic Peasible This alternative tain QueriedData Source Assessment Metrics User 7 Alternative7 This alternative contains DS1, DS2 and DS3 Assessment Metrics Dynamic Peasible This alternative tains that found in QueriedData Source Assessment Metrics User 7 Alternative7 This alternative contains DS1, DS2 and DS3 Assessment Metrics Dynamic Peasitive This alternative tains found in QueriedData Source Assessment Metrics User 7 Alternative7 This alternative contains DS1, DS2 and DS3 Assessment Metrics Dynamic Peasitive tains found in QueriedData Source Assessment Metrics User 2 0.93 0.93 0.93 0.92 3 0.64 0.93 0.92 3 0.44 0.33 0.34 (Con appletincess <td>4</td> <td>Alternative4</td> <td>This alternative</td> <td>Dynamic</td> <td>Peaktive</td> <td></td> <td>User</td>	4	Alternative4	This alternative	Dynamic	Peaktive		User
Assessment Metrics by DS1 and DS2 Assessment Metrics that found in QueriedData Assessment Metrics that found in QueriedData Assessment Metrics 3 Alternative3 This alternative centain DS3 Assessment Metrics Dynamic Pesitive This alternative centain DS3 Assessment Metrics User 6 Alternative3 This alternative centain DS3 Assessment Metrics Dynamic Pesitive This alternative centain DS3 Assessment Metrics User 6 Alternative3 This alternative centain DS3 Assessment Metrics Dynamic Pesitive This alternative centain DS3 Assessment Metrics User 7 Alternative7 This alternative centain DS3 Assessment Metrics Dynamic Pesitive This alternative centain DS3 Assessment Metrics User 7 Alternative7 This alternative centain DS3 Assessment Metrics Dynamic Pesitive centain DS3 Assessment Metrics User 7 Alternative7 This alternative centain DS3 Assessment Metrics Dynamic Pesitive centain DS3 Assessment Metrics User 7 Alternative7 This alternative centain Secure 3 Alternative centain DS3 Assessment Metrics Dynamic Pesitive centain DS3 Contain Metric Alternative centain DS3 Contain DS3 Contain DS3 Contain DS3 Contain DS3 DS4	-						
Alternative? This alternative contains DS1 and DS3 Assessment Metric Metric entity User and be realised on the second of the		1	Assessment Metrics			by DS1 and DS2	
5 Alternative 3 This alternative contains D31 and D33 Assessment Metrics Dynamic Pesitive This alternative dependence on the set of th		1					
S Alternative3 This elternative contains DS1 and DS3 Assessment Metric Dynamic Pesitive This elternative can be evaluated by DS3 Assessment Metric User 6 Alternative3 This elternative contains DS2 and DS3 Assessment Metric Dynamic Pesitive This elternative can be evaluated by DS3 assessment Metric User 6 Alternative3 This elternative contains DS2 and DS3 Assessment Metric Dynamic Pesitive This elternative can be evaluated by DS3 assessment Metric User 7 Alternative3 This elternative contains DS1, DS2 and DS3 Assessment Metric Dynamic Pesitive This elternative can be evaluated by DS3 adaption that found in QueriedData Sares ment Metric 7 Alternative3 This elternative contains DS1, DS2 and DS3 Assessment Metric Dynamic Pesitive This elternative contains DS1, DS2 and DS3 Assessment Metric 7 Alternative3 This elternative contains DS1, DS2 and DS3 Assessment Metric DS3 Assessment Metric This elternative contains Bare 7 Alternative3 Alternative Alternative contains Alternative Alternative Contains Alternative Alternative Contains Alternative Contains 7 Alternative Contains Alternative Contains Alternative Contains Alternative Contains Alternative Contains 8 Conseptences Alternative Contains Conseptences Con					1		
S Alternative 3 This alternative centains DS1 and DS3 Assessment Metrics Dynamic Pesitive Assessment Metrics User example of the valuated by DS1 and DS3 Assessment Metrics 6 Alternative 6 This alternative centains DS2 and DS3 Assessment Metrics Dynamic Pesitive This alternative centains DS1 assessment Metrics 6 Alternative 6 This alternative centains DS2 and DS3 Assessment Metrics Dynamic Pesitive This alternative centains DS2 and DS3 Assessment Metrics 7 Alternative 7 This alternative centains DS1, DS2 and DS3 Assessment Metrics Dynamic Pesitive This alternative centains DS1, DS2 and DS3 Assessment Metrics 7 Alternative 7 This alternative centains DS1, DS2 and DS3 Assessment Metrics Dynamic Pesitive This alternative centains DS1, DS2 and DS3 Assessment Metrics 7 Alternative 7 This alternative centains DS1, DS2 and DS3 Assessment Metrics Dynamic Pesitive This alternative centains DS1, DS2 and DS3 Assessment Metrics 8 Completiones Alternative Validity Alternative Alternative Centains DS2 O.93 O.93 8 O.93 0.93 0.92 O.92 O.92 4 0.38 0.34 0.34 O.34 8 0.34 0.38 0.34					1		
Alternative This alternative centains DS1 and DS3Assessment Matrice Dynamic Pesitive This alternative and be valuated by DS1 and DS3 Assessment Matrice User can be evaluated by DS1 and DS3 Assessment Matrice that found in Queries Data User 6 Alternative Contains DS2 and DS3Assessment Matrice Dynamic Pesitive This alternative can be evaluated by DS3 and DS3Assessment Matrice User 6 Alternative Contains DS2 and DS3Assessment Matrice Dynamic Pesitive This alternative can be evaluated by DS3 and DS3 Assessment Matrice that found in QueriedData Seure can be evaluated DS3Assessment Matrice User 7 Alternative? This alternative contains DS1, DS2 and DS3Assessment Matrice Dynamic Pesitive This alternative can be evaluated by DS3 assessment Matrice that found in QueriedData Seure Assessment Matrice that found in QueriedData Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment Matrice Seure Assessment					1		
3 Alternative 3 This alternative centation DS1 and DS3 Assessment Metrics Dynamic Pesitive This alternative centation DS3 Assessment Metrics User 6 Alternative6 This alternative centation Dynamic Pesitive This alternative assessment User 6 Alternative6 This alternative centation Dynamic Pesitive This alternative assessment User 6 Alternative6 This alternative centation Dynamic Pesitive This alternative assessment User 6 Alternative6 This alternative centation Dynamic Pesitive This alternative assessment User 7 Alternative7 This alternative centation Dynamic Pesitive This alternative assessment User 7 Alternative7 This alternative centation Dynamic Pesitive This alternative assessment User 7 Alternative7 This alternative Dynamic Pesitive This alternative assessment User 8 Alternative7 This alternative Dynamic Pesitive This alternative assessment Matrice that found in Queried Data Sacessment 3 0.34 0.32 0.32 0.32 0.32 0.32 4 0.33 0					1		
eentains DSI and DSIAssessment Metrics etn be evaluated by DSI and DSI Assessment Metrics etn be evaluated by DSI and DSI Assessment Metrics 6 Alternatived This alternative contains Dynamic Feasilive and DSI and DSI Assessment Metrics User 6 Alternatived This alternative contains Dynamic Feasilive and DSI and DSI Assessment Metrics User 7 Alternative? This alternative contains Dynamic Peasilive DSI and DSI Assessment Metrics User 7 Alternative? This alternative contains Dynamic Peasilive DSI and DSI Assessment Metrics User 7 Alternative? This alternative contains Dynamic Peasilive Table This alternative DSI and DSI Assessment Metrics User 4 Alternative Validity Assessment Metrics Alternative accurative accurative DSI and DSI Assessment Metrics Alternative accurati	5	Alternative5	This alternative	Dynamic	Pesitive	This alternative	User
6 Alternative This alternative Dynamic Pesitive This Joint Control Data 6 Alternative This alternative Dynamic Pesitive This User 6 Alternative District Dynamic Pesitive This User 6 Alternative District Dynamic Pesitive This User 7 Alternative? This alternative Dynamic Pesitive This User 7 Alternative? This alternative Dynamic Pesitive This Secure 7 Alternative? This alternative Dynamic Pesitive This Secure 7 Alternative? This alternative Assessment Metric District This Secure Assessment Metric District This Secure Assessment Metric entity Assessment Metric Secure Assessment Metric Secure As							
6 Alternative contains DS2 and DS3.Assessment Metric Dynamic Dynamic Pesitive This alternative cas be evaluated by DS3 and DS3. Assessment Metric User the evaluated by DS3.assessment Metric 7 Alternative? This alternative contains DS1, DS2 and DS3.Assessment Metric Dynamic Pesitive the evaluated by DS3.assessment Metric User the evaluated by DS3.assessment Metric Alternative Aggregated Metricid Alternative Validity the evaluated by DS3.assessment Metric Alternative the evaluated by DS3.assessment Metric Alternative the evaluated by DS3.assessment Metric Alternative Aggregated Tast Completences Alternative the output the o			DS3 Assessment Metrics		1		
6 Alternative of contains D32 and D53 Assessment Metrics Dynamic Penitive main Metric entity User was assessed on the second data and the					1		
6 Alternative Dynamic Feature Assessment Metric entry een tains User 6 Alternative This een tains DSJ and DSJ Assessment Metrice Dynamic Feature entry This elsemative is Queries User 7 Alternative? This entry alternative een tains Dynamic Feature is Queries This elsemative entry User 7 Alternative? This entry alternative een tains Dynamic Feature is Queries This elsemative entry User 7 Alternative? This entry alternative contains Dynamic Feature is Queries This elsemative entry User 8 Alternative Alternative Alternative entry This elsemative User 4 0.53 0.54 0.54 0.52 3 0.44 0.33 0.33 0.34					1		
6 Alternative This alternative central DS2 and DS3 Assessment Metrics Dynamic Pasitive This alternative central dots User 6 Alternative This alternative central dots DS3 Assessment Metrics Dynamic Pasitive This alternative can be valuated by DS2 and DS3 Assessment Metrics User Can be valuated by DS2 and DS3 Assessment Metrics Dynamic Pasitive This alternative can be valuated by DS2 and DS3 Assessment Metrics Dynamic Pasitive This alternative can be valuated by DS2 and DS3 Assessment Metrics Dynamic Pasitive This alternative can be valuated by DS3 and DS3 Assessment Metrics Dynamic Pasitive This alternative can be valuated by DS1, DS2 and DS3 Assessment Metrics Dynamic Pasitive This alternative can be valuated by DS1, DS2 and DS3 Assessment Metrics DS3 Assessment Metrics User Can by DS1, DS2 and DS3 Assessment Metrics <					1		
Alternative Metric entry 6 Alternative This alternative contains DS2 and DS3.Assessment Metrics Dynamic Fasilive and by DS3 and DS3 Assessment Metrics User 7 Alternative? This alternative contains DS1, DS2 and DS3.Assessment Metrics Dynamic Pasilive Metrics User 7 Alternative? This alternative contains DS1, DS2 and DS3.Assessment Metrics Dynamic Pasilive Metrics User 7 Alternative? This alternative contains DS1, DS2 and DS3.Assessment Metrics Dynamic Pasilive Tais alternative can be evaluated by DS3, Assessment Metrics User Alternative? Alternative Validity Assessment Metrics Alternative Can apple to the source of the passility User Alternative Alternative Validity Assessment Metrics Alternative Can be evaluated by DS3 Assessment Metrics Metrics that found in Queried Data Seurce Assessment Metrics Alternative Can apple to the Alternative Can be evaluated by DS3, DS3 Assessment Metrics 4 0.53 0.54 0.52 3 0.44 0.33 0.33 0.34 6 0.52 0.79 0.32					1		
6 Alternative of constainal DS2 and DS3Assessment Metrics Dynamic Pasitive This alternative constainal DS2 and DS3Assessment Metrics 7 Alternative? This alternative constainal DS2 and DS3Assessment Metrics Dynamic Dynamic Pasitive This alternative and DS3Assessment Metrics 7 Alternative? This alternative constainal DS1, DS2 and DS3Assessment Metrics Dynamic Dynamic Pasitive User 7 Alternative? This alternative constainal DS1, DS2 and DS3Assessment Metrics Dynamic Pasitive This alternative and DS3Assessment Metrics User 8 Alternative? This alternative Dynamic Pasitive This alternative and DS3Assessment Metrics User 8 Alternative? Alternative and DS3Assessment Metrics Dynamic Pasitive User 8 Alternative alternative and DS3Assessment Metrics Alternative and DS3Assessment Metrics Alternative alternative and DS3Assessment Metrics User 4 0.33 0.34 0.34 0.34 0.34 3 0.44 0.33 0.34 0.34					1		
een tains DS2 and DS3.Assessment n ean be evaluated by DS3 and DS3 Assessment notice Mathies Mathies that found in Querice/Data. Source Mathies ensity ? Alternative? This alternative contains Dynamic Peakinve This alternative User ? Alternative? This alternative contains Dynamic Peakinve This alternative User Alternative? Alternative Validity Alternative Accuracy Alternative Consument Mathie ensity User Alternative Alternative Validity Alternative Accuracy Alternative Consument Mathie ensity User Alternative Alternative Validity Alternative Accuracy Alternative Consument Mathie ensity Alternative Consument Mathie ensity 2 0.93 0.93 0.93 0.93 0.92 3 0.44 0.33 0.33 0.34 6 0.52 0.79 0.32	6	Alternative6	This alternative	Dynamic	Pesitive		User
Alternative 7 This alternative contains DS1, DS2 and DS3 Assessment Metrics Dynamic Pesitive This alternative contains DS1, DS2 and DS3 Assessment Metrics 7 Alternative 7 This alternative contains DS1, DS2 and DS3 Assessment Metrics Dynamic Pesitive This alternative contains to be a set of the contains of the contains contain				-	1	can be evaluated	
Alternative Alternative Dynamic Pesitive Metric shar found in Queriel Data Seurce Assessment Metric entity User 7 Alternative? This alternative contains DS1, DS2 and DS3 Assessment Metrice Dynamic Pesitive This alternative can be evaluated by DS1, DS2 and DS3 Assessment Metrice User Alternative Aggregated Aggregated Aggregative A			DS3 Assessment Metrics		1		
Alternative 7 This alternative contains D31, D52 and D53 Assessment Metrics Dynamic Penitive and the entity User was and the output of the entity 7 Alternative 7 This alternative contains D31, D52 and D53 Assessment Metrics Dynamic Penitive and the entity User was and D53 Assessment Metrics Alternative 4 Alternative 7 Alternative 7 Alternative 7 Alternative 7 Alternative 3 Alternative 7 Alternative 7 Alternative 7 Alternative 3 Alternative 7 Alternative 7 Alternative 3 0.33 0.32 3 0.44 0.38 0.38 6 0.52 0.79 0.32					1		
Alternative Alternative Dynamic Featility 7 Alternative? This alternative contains DS1, DS2 and DS3.Assessment Metrics Dynamic Featility This alternative can be evaluated by DS1, DS2 and DS3.Assessment Metrics User Alternative Alternative Alternative User Alternative Alternative Alternative User Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative 3 0.63 0.62 0.67 3 0.54 0.54 0.52 5 0.44 0.33 0.33 6 0.52 0.79 0.32					1		
Alsonative? This alsonative cention Assessment Metrice mity ? Alsonative? This alsonative centain DS1, DS2 and DS3 Assessment Metrice Dynamic Pesitive This alsonative centain also centain DS1, DS2 and DS3 Assessment Metrice Alsonative? Alsonative? Dynamic Pesitive Pesitive This alsonative centain also also centain also c					1		
Alternative? This alternative contains DS1, DS2 and DS3.Assessment Metrice Dynamic Pasitive Status This alternative entains DS1, DS2 and DS3.Assessment Metrice User Alternative? Alternative Validity Assessment Metrice Assessment Metrice User Alternative? Alternative Validity Alternative Metrice that found in Queried Data Seurce Assessment Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Assessment Metrice Alternative Alternative Alternative Alternative Alternative Alternative Assessment Metrice 0.93 0.93 0.92 3 0.68 0.62 0.67 3 0.44 0.38 0.33 0.34 6 0.52 0.79 0.92					1		
Alternative Alternative Validity Alternative Alternative Aggregated Fact Alternative Validity Alternative 2 0.33 0.33 0.33 3 0.68 0.62 0.67 3 0.48 0.33 0.34 4 0.33 0.33 0.34 6 0.32 0.79 0.32					1		
Alternative Alternative Alternative Alternative Aggregated Fact Alternative Alternative 2 0.33 0.93 0.93 3 0.68 0.62 0.67 3 0.53 0.54 0.54 5 0.44 0.38 0.33 6 0.52 0.79 0.92	7	Alternative/	This alternative	Dynamic	Pesitive	This alternative	User
Alternative Alternative Validity Alternative e Validity Alternative Alternative Validity Aggregated Fact Metricid Campletineces Metricid 0.93 0.03 0.93 3 0.68 0.54 0.54 3 0.44 0.33 0.33 0.33 0.34 0.33 0.32		1		-			
Alternative Alternative Alternative Alternative Alternative Alternative Alternative Alternative Aggregated Fact Alternative Alternative Metridd Completeness Completeness Metric mity 2 0.93 0.93 0.93 0.92 3 0.68 0.62 0.62 0.67 3 0.44 0.38 0.38 0.34 6 0.32 0.79 0.32		1	DS3 Assessment Metrics				
Alternative Alternative Validity Alternative Control Data		1					
Alternative Alternative Alternative Alternative Source Assessment Aggregated Metrical 2 Fact Completeness Alternative Alternative Accuracy Alternative Timeliness 2 0.93 0.93 0.93 0.92 3 0.68 0.62 0.62 0.67 3 0.44 0.38 0.34 0.34 6 0.32 0.79 0.79 0.92		1					
Alternative Aggregated Alternative Validity Factor Alternative Validity Alternative Communication Alternative Communication Alternative Communication Alternative Communication 3 0.93 0.93 0.93 0.92 4 0.53 0.54 0.92 0.84 5 0.44 0.38 0.33 0.84 6 0.52 0.79 0.79 0.92		1					
Alternative Aggregated Metricid Alternative Fact Alternative Completences Alternat		1					
Alternative Aggregated Metridd Alternative Fact Alternative Aggregates Alternative Accuracy Alternative Timelines 2 0.93 0.95 0.92 3 0.68 0.62 0.62 4 0.53 0.54 0.52 5 0.44 0.53 0.53 0.54 6 0.52 0.79 0.79 0.79		1				Assessment	
Argregated MetricId Fact Completances Accuracy Timeliness 2 0.95 0.95 0.95 0.92 3 0.68 0.62 0.62 0.67 4 0.53 0.54 0.34 0.92 5 0.44 0.38 0.38 0.34 6 0.52 0.79 0.79 0.79						Metric entity	
Microid Complexences 0.93 0.93 0.92 2 0.93 0.93 0.93 0.92 3 0.68 0.62 0.62 0.67 4 0.58 0.54 0.54 0.92 5 0.44 0.58 0.54 0.52 6 0.52 0.79 0.79 0.92			Alternative Validity				
2 0.95 0.93 0.93 0.92 3 0.68 0.62 0.62 0.67 4 0.53 0.54 0.54 0.92 5 0.44 0.38 0.38 0.34 6 0.32 0.79 0.79 0.92				Accuracy	Timeliness		
3 0.63 0.62 0.62 0.67 4 0.58 0.54 0.54 0.92 5 0.44 0.33 0.33 0.34 6 0.32 0.79 0.79 0.92			0.95	0.05	0.92		
4 0.38 0.34 0.34 0.32 5 0.44 0.38 0.38 0.38 0.34 6 0.32 0.79 0.79 0.92							
5 0.44 0.38 0.38 0.34 6 0.82 0.79 0.79 0.92							
6 0.82 0.79 0.79 0.92			0.38		0.84		
7 0.61 0.54 0.57 0.92	6	0.82	0.79		0.92		
	7	0.61	0.54	0.57	0.92		

According to the required quality level in Q2, we will prune Alternative4, Alternative5 and Alternative7 as they are under the required level of quality and the final qualified alternatives metrics for Q2 will be presented in table 26.

Table 26.AlternativeAggregatedMetric entity for Q2 after second pruning

Alternative Aggregated MetricId	Alternative Aggregated MetricName	Alternative Aggregated MetricDefinition	Alternative Aggregated MetricTemporary	Alternative Aggregated MetricState	Metric History	Information Supplier
2	Alternative2	This alternative contains DS2Assessment Metric and it can satisfy the required level of data quality where query answer completeness≥0.65	Dynamic	Positive	This alternative evaluated by Ds2 Assessment Metric that found in QueriedData Source Assessment Metric entity	User
3	Alternative3	This alternative contains DS3Assessment Metric and it can satisfy the required level of data quality where query answer completeness≥0.65	Dynamic	Positive	This alternative evaluated by Da3 Assessment Metric that found in QueriedData Source Assessment Metric entity	User
6	Alternative6	This alternative contains DS2 and DS3Assessment Metrics and it can satisfy the required where query answer completeness≥0.65	Dynamic	Positive	This alternative can be evaluated by DS2 and Ds3 Assessment Metric that found in QueriedData Source Assessment Metric entity	User
Alternative Aggregated MetricId	Alternative Fact Completeness	Alternative Validity	Alternative Accuracy	Alternati Timeline		
2	0.95	0.95	0.95	0.92		
3	0.68	0.62	0.62	0.67		
6	0.82	0.79	0.79	0.92		

3.4.5 Alternatives Ranking

For many years, the advantages of databases and information retrieval systems have merged to achieve the goal of many researchers. While database systems provide efficient treatment with data, mechanisms for effective retrieval and fuzzy ranking[8] that are more attractive to the user are provided with IR. In our work, we will rank alternatives according to their data quality features scores and according to different queries types

3.4.5.1 Ranking Alternatives according to Proposed Queries Types

We will present different queries' types depending on number of quality features in query condition (from one to four) and the kind of quality features' value (quantitative or qualitative). Quantitative data are values presented as numbers and qualitative data are values presented by a name, symbol, or a number code and they require user intervention as low, medium and high as features' values can represent different scores to different users.

I. No-Feature Top-K Selection Query

In this type, queries don't include any specified quality condition, so we will build AlternativeAggregatedMetric table 28 from QueriedDataSourceAssessmentMetric table 27. Then, we will return to user all alternatives ranked according to all proposed features as presented in table 29, table 30, table 31, and table 32 and we will let him to choose the most suitable feature ranking

User Query (Q3)	Q3: Select SName, SAddress, DOB, SupName From G Limit 3
Cati Query (Q3)	

Table 27.QueriedDataSourceAssessmentMetric entity for Q3

QueriedDS Assessment MetricId	Queried Data Source	Asses	ridDS sment cName	QueriedDS Assessment MetricDefinition	QueriedDS Assessment Metric	QueriedDS Assessment MetricState	QueriedDS Assessment Metric
1	<u>Id</u> 1	DS2Assessment Metric		This metric is evaluation for DS2 required attributes data quality	Temporary Dynamic	Positive	History This evaluation metric is for DS2 that found in QueriedDataSource entity with its required attributes
2	2	DS3Assessment Metric		This metric is evaluation for DS3 required attributes data quality	Dynamic	Positive	This evaluation metric is for DS3 that found in QueriedDataSource entity with its required attributes.
		QueriedDS Assessment Metric Id	Information Supplier	QueriedDS FactCompleteness	QueriedDS Validity	QueriedDS Accuracy	QueriedDS Timeliness
		1 2	User User	0.95 0.68	0.95 0.62	0.95 0.62	0.92 0.67

Table 28.AlternativeAggregatedMetric entity for Q3

International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.6, No.2, March 2016

Alternative	Alternative	Alternative	Alternative	Alternative	Alternative	Information
Aggregated MetricId	Alternative Aggregated MetricName	Alternative Aggregated MetricDefinition	Alternative Aggregated MetricTemporary	Aggregated MetricState	Aggregated Metric	Supplier
1	Alternativel	This alternative	Dynamic	Positive	This	User
1	Alternativel	contains	Dynamic	Positive	alternative	User
		DS2Assessment			can be	
		Metric			evaluated by	
					Da2	
					Assessment	
					Metric that found in	
					QueriedData	
					Source	
					Assessment	
					Metric entity	
2	Alternative2	This alternative	Dynamic	Positive	This	User
		contains DS3Assessment			alternative can be	
		Metric			evaluated by	
		ivietiic.			Da3	
					Assessment	
					Metric that	
					found in	
					QueriedData Source	
					Assessment	
					Metric entity	
3	Alternative3	This alternative	Dynamic	Positive	This	User
		contains D2 and			alternative	
		DS3Assessment			can be	
		Metrics			evaluated by DS2 and Ds3	
					Assessment	
					Metric that	
					found in	
					QueriedData	
					Source	
					Assessment Metric entity	
Alternative	Alternative	Alternative	Alternative		intente carry	
Fact	Validity	Accuracy	Timeliness			
Completeness 0.95	0.95	0.95	0.92	-		
0.95	0.95	0.95	0.92	-		
0.82	0.79	0.79	0.92	-		
	Table 29. Top 3			able 30.Top 3		
$\langle \cdot \rangle$	Ranked acc	ording to		able 30.Top 3 nked accordin		
		ording to				
	Ranked acc	ording to	Ra	nked accordin	ng to Validity	
Alternative	Ranked acc FactComp	ording to leteness	Alterna	anked accordin	AlternativeVa	didity
	Ranked acc FactComp	ording to leteness	Ra	anked accordin	AlternativeVa	didity
	Ranked acc FactComp	Alternative ctCompleteness 0.95	Alterna	ativeName	AlternativeVa 0.95 0.79	llidity
Alternative	Ranked acc FactComp Name Fa 1 3	Alternative ctCompleteness	Alterna	ativeName ative1 ative3	AlternativeVa	llidity
Alternative	Ranked acc FactComp Name Fa 1 3	Alternative ctCompleteness 0.95	Alterna Alterna Alterna	ativeName ative1 ative3	AlternativeVa 0.95 0.79	didity
Alternative	Ranked acc FactComp Name Fa 1 3	Alternative ctCompleteness 0.95 0.82	Alterna Alterna Alterna	ativeName ative1 ative3	AlternativeVa 0.95 0.79	didity
Alternative Alternative Alternative	Ranked acc FactComp	Alternative ctCompleteness 0.95 0.82 0.68	Alterna Alterna Alterna	ativeName ative1 ative2	AlternativeVa 0.95 0.79	didity
Alternative Alternative Alternative	Ranked acc FactComp Name Fa 1 3	Alternative ctCompleteness 0.95 0.82 0.68	Alterna Alterna Alterna	ativeName ativeName ttive1 ttive2 Table 32.Top	AlternativeVa 0.95 0.79 0.62	
Alternative Alternative Alternative	Ranked acc FactComp	Alternative ctCompleteness 0.95 0.82 0.68	Alterna Alterna Alterna	ativeName ativeName ttive1 ttive2 Table 32.Top	AlternativeVa 0.95 0.79 0.62 3 Alternatives	
Alternative Alternative Alternative	Ranked acc FactComp	Alternative ctCompleteness 0.95 0.82 0.68 Alternatives	Alterna Alterna Alterna	ativeName ativeName ttive1 ttive2 Table 32.Top	AlternativeVa 0.95 0.79 0.62 3 Alternatives	
Alternative Alternative Alternative R	Ranked acc FactComp	Alternative ccCompleteness 0.95 0.82 0.68 Alternatives g to Accuracy	Alterna Alterna Alterna Alterna Alterna	tiveName tive1 tive2 Table 32.Top tanked accord	AlternativeVa 0.95 0.79 0.62 3 Alternatives ing to Timelines	s
Alternative Alternative Alternative R Alternative	Ranked acc FactComp	Alternative ctCompleteness 0.95 0.82 0.68 Alternatives ag to Accuracy	Alterna Alterna Alterna Alterna Alterna Cy Alter	nnked accordin ativeName ttive1 ttive2 Table 32. Top tanked accord nativeName	Alternatives 0.95 0.79 0.62 3 Alternatives ing to Timelines AlternativeTh	s
Alternative Alternative Alternative R Alternative	Ranked acc FactComp	Alternative ccCompleteness 0.95 0.82 0.68 Alternatives g to Accuracy AlternativeAccura 0.95	Alterna Alterna Alterna Alterna Alterna Cy Altern Alterna	naked accordin ativeName tive1 tive2 Table 32. Top tanked accord nativeName native1	AlternativeVa 0.95 0.79 0.62 3 Alternatives ing to Timelines AlternativeTin 0.92	s
Alternative Alternative Alternative R Alternative	Ranked acc FactComp	Alternative ctCompleteness 0.95 0.82 0.68 Alternatives ag to Accuracy	CY Altern Alterna Alterna Alterna Alterna CY Alterna Alterna	nnked accordin ativeName ttive1 ttive2 Table 32. Top tanked accord nativeName	Alternatives 0.95 0.79 0.62 3 Alternatives ing to Timelines AlternativeTh	s

I. Single-Feature Top-K Selection Query

In this type, queries include one data quality feature as a quality condition, so we will build AlternativeAggregatedMetric table from QueriedDataSourceAssessmentMetrice table but after first and second pruning according to specified quality condition in user query and then return to user alternatives ranked according to required data quality feature in user query.

This type of queries classified into two categories Quantitate Single-Feature Top-K Selection Query and Qualified Single-Feature Top-K Selection Query. We will introduce every category as following:

• Quantitate Single-Feature Top-K Selection Query

In Quantitate Single-Feature Top-K Selection Query, Data quality features values are presented as quantitate values.

A SQL template for Single-Feature top-k selection query is the following:

SELECT some attributes FROM G WHERE selection condition WITH data quality feature condition ORDER BY F $(P_1, \ldots P_m)$ LIMIT k

A SQL example for Quantitate Single-Feature top-k selection query (Q4) is the following:

Select SName, SAddress, DOB, SupName, DName From G With AlternativeFactCompleteness≥ 0.65 Order by AlternativeFactCompleteness desc Limit 3

• Qualitative Single-Feature Top-K Selection Query

In Qualitative Single-Feature Top-K Selection Query, Data quality features values are represented as qualitative values.

A SQL example for Qualified Single-Feature top-k selection query (Q5) is the following:

	Select SName, SAddress, DOB, SupName, DName
	From G
	With AlternativeFactCompleteness is high
	Order by AlternativeFactCompleteness desc
	Limit 3
-	

Received user message:

High represents AlternativeFactCompleteness ≥ 0.65

Using the AlternativeAggregatedMetric table 26 that satisfies (Q4 and Q5), the ranked alternatives are presented in table 33

Table 33. Top 3 Alternatives Ranked according to FactCompleteness							
Alternative	1						
FactCompleteness							
0.95							
0.82							
0.68							
	Alternative FactCompleteness 0.95 0.82						

II. Multi- Feature Top-K Selection Query

In this type, queries include many data quality features as a condition, so we will build AlternativeAggregatedMetric table from QueriedDataSourceAssessmentMetric table but after first and second pruning according to specified quality condition in user query and then return to user alternatives ranked according to user query case.

Case1: Data quality features specified in user's query are separated with (AND) and all are satisfied.

In this case, we will consider queried data source or alternative as qualified one if it satisfies all required data quality features together. The AlternativesAggregatedMetric table will be built from qualified alternatives and they will be ranked according to total score by TA algorithm.

Case2: Data quality features specified in user's query are separated with (AND) and the required level of quality for one or more of data quality features doesn't commensurate with the required level of quality for other data quality features or doesn't achieve.

In this case, there are no queried data sources or alternatives can return required query attributes with specified quality levels so, message will be sent to user to inform him that his required level of quality for query answering can't be satisfied with these data quality features together.

Case3: Data quality features specified in user's query are separated with (OR) and the required level of quality for all data features satisfied or the required level of quality for one or more of data quality features doesn't commensurate with the required level of quality for others data quality features or can't be achieved.

In this case, we will consider queried data source or alternative as qualified one if it satisfies at least one required data quality feature. The AlternativesAggregatedMetric table will be built from qualified alternatives and they will be ranked according to total score by TA algorithm.

This type of queries classified into two categories Quantitate Multi-Feature Top-K Selection Query and Qualified Multi-Feature Top-K Selection Query. We will introduce every category as following:

• Quantitate Multi-Feature Top-K Selection Query

In this type queries condition contains multi-features and their values are represented in quantitate way.

A SQL template for multi-Feature top-k selection query is the following:

```
SELECT some attributes
FROM G
WHERE selection condition
WITH data quality feature condition
ORDER BY F (P_1, \dots P_m)
LIMIT k
```

A SQL example for Quantitate Multi-Feature top-k selection query (Q6) in (Case1) is the following:

Select SName, SAddress, DOB, SupName, DName
From G
With AlternativeFactCompleteness ≥ 0.65 and AlternativeValidity ≥ 0.65 and
AlternativeAccuracy ≥ 0.65
Order by AlternativeFactCompleteness desc, AlternativeValidity desc,
AlternativeAccuracy desc
Limit 3

• Qualified Multi-Feature Top-K Selection Query

In this type of queries, the query condition contains more than one feature but the values of data quality features are represented in qualitative way.

A SQL example for Qualified Multi-Feature top-k selection query (Q7) (Case1) is the following:

Select SName, SAddress, DOB, SupName, DName	
From G	
With AlternativeFactCompleteness high and Alternativ	eValidity high and
AlternativeAccuracy high	
Order by AlternativeFactCompleteness desc, Alternativ	eValidity desc,
AlternativeAccuracy desc	
Limit 3	

Received User message:

High Validity represents AlternativeValidity ≥ 0.65 , High Fact Completeness represents AlternativeFactCompleteness ≥ 0.65 and High Accuracy represents AlternativeAccuracy ≥ 0.65

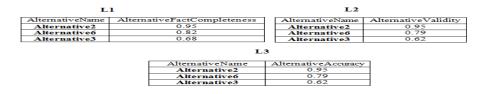
To deal with Multi-Features Top-K Selection Queries, we should build Lists (tables); every list contains alternatives from alternative aggregated metric and it ranks descending according to one of data quality features' scores that is required in user query, so we need to combine these ranking to produce global ranking.[9]

In our work, we choose **Threshold Algorithm (TA)** proposed by Fagin et.al.2001 as ranking algorithm[10]. It considers famous, simple and elegant Top-K algorithm, it considers the basic algorithm for all next variants, it is applicable for queries where the scoring function is monotonic, it is based on an early-termination condition and it evaluates top-k queries without examining all the tuples. This algorithm is presented as following:

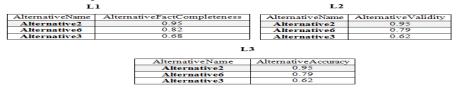
Algorithm: TA [Fagin et al. 2001]
(1) Do sorted access in parallel to each of the m sorted lists L_i . As a new object (o) is seen
under sorted access in some list, do random access to the other lists to find P_i (o) in every
other list L_i . Compute the score $F(o) = F(P_1, \ldots, P_m)$ of object o. If this score is among the k highest scores seen so far, then remember object o and its score $F(o)$.
(2) For each list L_i , let P_i be the score of the last object seen under sorted access. Define the
threshold value T to be F (P_1, \ldots, P_m) . As soon as at least k objects have been seen with scores at least equal to T, halt.
(3) Let A_k be a set containing the k seen objects with the highest scores. The output is the sorted set {(o, F (o)) $ o \in A_k$ }.

Figure 3.Threshold algorithm (TA)[11]

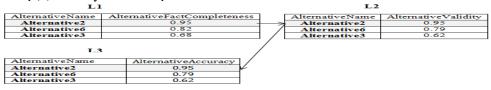
Using the AlternativeAggregatedMetric table 26 that achieves (Q6 and Q7) and applying TA algorithm to following lists, the ranked alternatives will be presented in table 36



1) Sorted access in parallel to each of the 3 sorted lists.



2) For every new object o is seen under sorted access in some list, do random access to the other lists to find P_i (o) in every other list L_i .

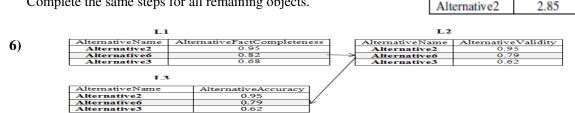


- 3) Compute the score $F(o) = F(P_1, ..., P_m)$ of object o. If this score is among the k highest scores seen so far, then remember object o and its score F (o). Assume that ranking Function is sum, so F (Alternative 2) = 2.85
- 4) The threshold value T is $F(P_1, \ldots, P_m)$ for the scores of the last seen object The threshold value T = 2.85**Because** F (Alternative2) = the threshold value T = 2.85 so, Alternative2 will put in A_k .
- 5) A_k Is a set containing the k seen objects with the highest scores

Table 34. Ak with k seen objects with the highest scores

F(O)

Complete the same steps for all remaining objects.



7) F (Alternative6) = 2.4

8) The threshold value T = 2.4

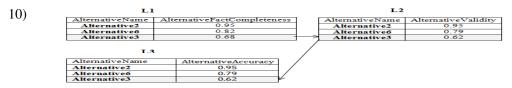
9) Because F (Alternative6) = the threshold value

T = 2.4 so, Alternative6 will put in A_k and rank the existing alternatives in descending order in

	objects with the highest scores		
A_k	0	F(O)	
	Alternative2	2.85	
	Alternative6	2.4	

Table 35.Ak with k seen

0



11) F (Alternative3) = 1.92

- 12) The threshold value T = 1.92
- 13) **Because** F (Alternative3) = the threshold value T = 1.92 so, Alternative3 will put in A_k . Alternative2, Alternative6 and Alternative3 will put in descending order in A_k

 A_k

In our example, the alternatives ranking is the same for all data quality features that specified in user query so we can directly say that top-3 ranking alternatives that satisfy user requirements from quality are Alternative2, Alternative6 and Alternative3.

Table 36.Ak with k seen objects with the highest scores

0	F(O)
Alternative2	2.85
Alternative6	2.4
Alternative3	1.92

Duplicate Detection and Data Fusion Process

Some of Top-K Ranked Alternatives produced by DIRA consist of one qualified queried data source and others consist of more than one queried data source, those containing more than one queried data source will pass to duplicate detection and data fusion algorithms that will run on their results respectively then these results will be re-evaluated using assessment metrics that will use equations 5, 12, 13 and 14 which are presented in section1 and section 3.2 to be added to alternatives with one qualified queried data source to re-rank to return final top-k alternatives.

4. CONCLUSION

In this paper, we presented data integration framework that integrates large number of available data sources with different levels of quality to return top-k qualified query answers from significant ones only.

This framework introduces new accurate type of completeness called fact-completeness that will be used in DIRA assessment module that works on four data quality features completeness, validity, accuracy and timeliness for early pruning of data sources under the required level of quality and retrieving data from only qualified ones, this framework also shortens processing time of duplicate detection and data fusion as they will work on only top-k alternatives with more than one queried data source not all available query results and it can be extended to include different types of data sources, add more data quality features and use different ranking algorithm.

REFERENCES

- [1] M. S. Abdel-Moneim, A. H. El-Bastawissy, and M. H. Kholief, "Quality Driven Approach for Data Integration Systems," 7th Int. Conf. Inf. Technol., vol. 2015, pp. 409–420, 2015.
- [2] M. Kaiser, "A Conceptual Approach to Unify Completeness, Consistency, and Accuracy as Quality Dimensions of Data Values," vol. 2010, pp. 1–17, 2010.
- [3] C. Moraes and A. C. Salgado, "Information Quality Measurement in Data Integration Schemas," ACM, 2007.
- [4] C. Batini and M. Scannapieco, Data Quality Concepts, Methodologies and Techniques. 2006.
- [5] P. N. Mendes, H. Mühleisen, and C. Bizer, "Sieve: Linked Data Quality Assessment and Fusion," Proc. 2012 Jt. EDBT/ICDT Work., pp. 116–123, 2012.
- [6] I. N. R. Etrieval, "A Flexible Quality Framework for Use Within Information Retrieval," Quality, pp. 297–313.
- [7] P. Angeles and F. García-ugalde, "A Data Quality Practical Approach," vol. 2, no. 2, pp. 259–274, 2009.
- [8] R. Fagin, A. Lotem, and M. Naor, "Optimal Aggregation Algorithms for Middleware," J. Comput. Syst. Sci., vol. 66, no. 4, pp. 614–656, 2003.
- [9] I. F. Ilyas, G. Beskales, and M. A. Soliman, "A Survey of Top-k Query Processing Techniques in Relational Database Systems," ACM Comput. Surv., vol. 40, no. 4, pp. 1–58, 2008.
- [10] G. Das, D. Gunopulos, N. Koudas, and D. Tsirogiannis, "Answering Top-k Queries Using Views," Proc. 32nd Int. Conf. Very Large Data Bases, pp. 451–462, 2006.
- [11] I. F. Ilyas, G. Beskales, and M. a. Soliman, "Query Processing Techniques in Relational Database Systems," ACM Comput. Surv., vol. 40, no. 4, pp. 1–58, 2008.