

Data Integration: Viewpoint Based Approach

Bouchra Boulkroun^{#1}, Fouzia Benchikha^{*2}

[#]LIRE Laboratory, Computer Sciences Department,
Constantine 2 University, Algeria

¹boulkroun_bouchra@hotmail.fr

^{*}LIRE Laboratory, Constantine 2 University, Algeria

Sciences Faculty, 20 August 1955 University of Skikda, Algeria

²f_benchikha@yahoo.fr

Abstract— Nowadays, the mediation of information from multiple autonomous and heterogeneous data sources becomes a crucial task for future Web information systems. With the succession of ontologies, an important progress has been realized in the automation of the integration process seen their explicit representation of data semantics. Somewhere else, the paradigm of viewpoints can be an undeniable contribution in the integration of large number of information sources by several viewpoints. In this paper we propose an hybrid approach of integration based on the notion of viewpoints which takes into account new hypotheses: (1) data sources are grouped according to the viewpoint that they represent, (2) mediators are represented in two different levels of the architecture, and (3) an hybrid approach that consists in combining the two data integration approaches Global As View (GAV) and Local As View (LAV) is used.

Keywords— Integration, viewpoints, mediation, ontologies, GAV, LAV.

I. INTRODUCTION

The exponential development of information exchange over the web augments the difficulties to find the relevant information desired by an end user. This information is often distributed on several heterogeneous data sources. Thus, the integration of such sources becomes more and more a crucial need. Several integration systems are proposed in the literature. They are classified in two categories: the first category concerns the most traditional systems based essentially on the structure of DBs, they represent the *multibases systems* [5], [16]. The second category is most known recently and concerns, both *virtual* and *materialized* approaches [8]. Several extensions of integration systems are proposed to improve the activity of integration. Moreover, in every case where many designers of data sources present various representations, we cannot eliminate the existence of several interpretations that must be integrated to represent the whole system and to answer any need on it.

In this context we propose the introduction of viewpoints (VP) paradigm for data integration. Our approach deals with the interpretations of every user implied in the domain. Mainly, it is based on the identification of viewpoints included in data sources. The approach proposed in this paper, is the result of the combination of the best of GAV and LAV data integration approaches, so reducing the problem of queries rewriting while being easily cover-up. Thus, our integration

process is defined on two essential levels: (i) a classification of data sources in categories according to various viewpoints, (ii) an integration of these last following the definitions in the domain ontology. The rest of this paper is organized as following: in section two we present a brief description of data integration and in section three we define the notion of viewpoints. In section four we present an outline of our approach followed by the description of the proposed architecture in section five. An example of query's treatment is detailed in section six. Section seven presents a comparison study followed by a conclusion in section eight.

II. DATA INTEGRATION

The data integration activity allows giving a transparent access to heterogeneous data sources by the use of an abstract common scheme. So, the main issue of integration systems is how to efficiently provide users with a flexible access to information distributed on multiple autonomous data sources [9]. These systems play the role of middleware between all the data sources and the users without caring about their structure or their location [3]. Two main approaches for integration are distinguished: (i) data warehouse approach (materialized approach), allowing physical data integration by transferring the data towards a single space of storage [8], and (ii) virtual approach (or mediation), which constitute the most common solution to integrate heterogeneous data sources [7].

The mediation architecture consists in developing an application asked to play the role of interface between the local data sources and the user's application. The mediation systems can differ according to the defined correspondence between the global scheme and data sources. Two classic approaches are distinguished: The first approach Global-As-View (GAV) [10], consists in defining the global scheme as a view on the local schemas of data sources. According to this approach any change in the local sources (the addition of a data source) implies the redefinition of the global scheme. This architecture represents the most used one; many systems are built under it (MOMIS [2]). On the other hand, the Local-As-View approach (LAV) [10] consists in defining the local schemas as views on the global scheme. The global scheme here is defined according to the needs of users. The rewriting of requests presents a more complex task and requires complicated algorithms. This approach is adopted in few systems such as (OBSERVER [12], PICSEL [14]).

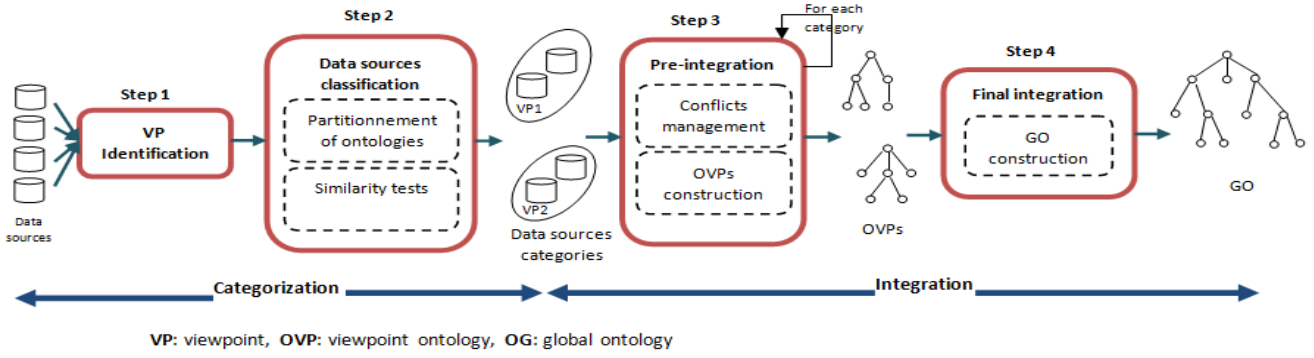


Fig. 1 Viewpoint-based integration approach

III. THE VIEWPOINT NOTION

Viewpoints are defined in several ways, according to their use in various domains of data processing. In the context of knowledge representation Nguyen and Rieu [7], defines the viewpoint as being: "synonym of an object representation which possesses several descriptions and is submitted to several users". According to Shilling and Sweemey [15], a viewpoint can be defined as: "a simplified abstraction of a complex structure by deleting non relevant information for the current objective". Benchikha and all [1] in their works on object oriented databases, suppose that the viewpoint-based approach must be based on the conjunction actor / information. They define the viewpoint as: "a conceptual means binding on the one hand the actor who observes and, on the other hand a phenomenon (or a world) which is observed".

In the context of our work the viewpoints definition relies three essential points: entity – actor – context or objective.

- Entity; the object that can be studied in a particular domain.
- Actor; the person who intervenes to study / observe a particular domain.
- Context or objective; a situation where the actor can see the domain entities.

We consider the following definition: "a viewpoint is a simplified abstraction of an object which can have several representations characterizing different contexts around which this object can be observed". In the next section, we will present our proposed approach using viewpoints concept.

IV. THE PROPOSED APPROACH

Our proposed approach (Cf. fig.1), is structured in two processes: categorization and integration.

A. Categorisation process

In this process we concentrate essentially on the study of viewpoints represented in the data sources to be integrated.

1) *Viewpoints Identification*: The identification of VPs described by different expert's results from a detailed study of data sources. To define a viewpoint implied by the data sources, we have to consider the part of local ontology in every source. A viewpoint is formally defined by the quadruplet: $VP : \langle C^{VP}, P^{VP}, Att^{VP}, Sub \rangle$, where:

- C^{VP} : represent all the concepts implied in the definition of an expert viewpoint.

- P^{VP} : represent all the valid properties according to the expert for the representation of a viewpoint.
- Att^{VP} : represent all the visible attributes¹ for a VP.
- Sub: $C \rightarrow 2^C$ is the relation which allows to define for every class c_i its subsumed classes.

2) Data Sources Classification

The objective of this activity is to build data sources categories which consider the different viewpoints of the experts. The classification process is represented below:

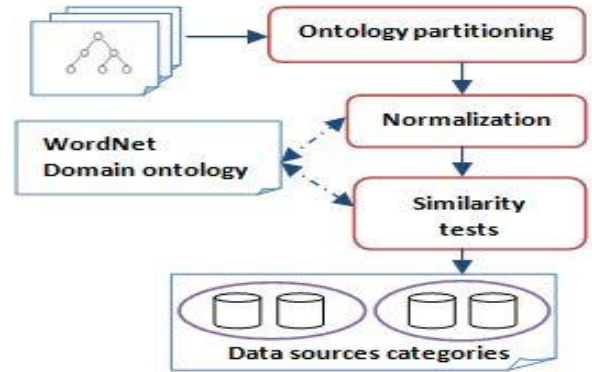


Fig. 2 Data sources classification process

a) *Ontology partitioning*: The objective of this phase is to divide ontology into blocks b_1, b_2, \dots, b_n . We start from the general concept in the hierarchy and we group them gradually in deeper blocks.

Algorithm: Ontology partitioning

Input : O (set of ontologies)

Begin

```

Initialize  $C_i$  in T; // T represents the most general concept
If subclass ( $C_i$ )  $\neq \emptyset$  do // when the concept  $C_i$  have a set of sub
classes they are added to the same block of the concept  $C_i$ 
  Add subclass ( $C_i$ ) in  $b_j$ ;
  For  $i = 1.. n$  do//browse the set of classes in  $b_j$ 
    If subClass (subclass( $C_i$ ))  $\neq \emptyset$  do
       $j := j+1$ ; // Create a new block for the sub classes of the
concept  $C_i$  and their sub classes

```

¹ For a particular VP only the relevant attributes for the latter are visible for a concept defined in the hierarchy

Add subclass (subclass (C_i)) to b_j .

End

Output : B (set of blocks after partitioning)

- b) *Normalization*: This phase consists in normalizing the names of classes used in the representation of viewpoints. The textual normalization of the names of classes concerns various forms such as: the normalization of characters (e.g. CarS: cars), the deletion of diacritical marks or accents (e.g. Hanoï: Hanoi), the deletion of numbers, the elimination of punctuation (e.g. specific_entity: {specific, entity}), the extension of abbreviations (e.g. SW: {Semantic, Web}), etc.
- c) *Similarity tests*: The activity of data sources classification proposed in our approach, can take advantage of alignment techniques to make a first comparison between the data sources to integrate them into VP ontologies. By inspiring from the work of Abdenour and Bruno [4] which used the similarity between VP by using the notion of roles between the human actor and the agent, and by taking advantage from some similarity measures already identified within the framework of ontologies alignment, we propose a model of similarity allowing to calculate lexical and structural similarity:

Concepts similarity:

- **Lexical similarity**: We propose a linguistic comparison method based on the thesaurus WordNet [8]. The WordNet similarity (WNSim) is defined as follow:

$$\text{WNSim}(C1, C2) = 1: \text{IF } (C1 \text{ and } C2 \text{ are synonyms}) \\ \text{Else } (\exists \text{ at least a means of a} \\ \text{terminological relation in WordNet} \\ \text{between the synsets of } C1 \text{ and } C2) \\ 0: \text{the opposite case} \quad (1)$$

- **Structural similarity**: The treatment of similarity in this stage is based on the concepts structure. The similarity between concepts is measured by comparing the informative contents shared between the latter. Here the notion of IC (Informative Content) is introduced [13]. The IC of a concept is introduced as follows:

$$\text{IC}(c) = -\log(P(c)) \quad (2)$$

Where $P(c)$, is the probability to find an instance of the concept c . This probability is calculated by:

$$P(c) = \text{freq}(c) / N \\ \text{freq}(c) = \sum_{n \in \text{word}(c)} \text{count}(n) \quad (3)$$

Where: $\text{word}(c)$ represents all the terms representing the concepts subsumed by c , and N is the total number of the instances of the concept c .

We adopted also the Lin measure [11] which considers also the informative contents between two concepts. The latter is formulated as follows:

$$\text{Sim}(c_1, c_2) = 2 * \log(P(c)) / \log(P(c_1)) * \log(P(c_2)) \quad (4)$$

The values of partial similarities obtained previously, have to be composed to form the global similarity between concepts and they are balanced by coefficients. The

coefficient associated with the value of lexical similarity (WNSim) for example, is given by the following formula:

$$\text{WN} = e^{\text{WNSim}} \quad (5)$$

$$\text{CW} = e^{\text{Sim}} \quad (6)$$

Other coefficients are calculated in the same way. Consequently, the global similarity (SimG) is given by:

$$\text{Sim}_G(c_1, c_2) = (\text{WN} * \text{WNSim} + \text{CW} * \text{Sim}) / (\text{LW} + \text{CW}) \quad (7)$$

Attribute similarity:

In this stage we calculate the similarity between the attributes of the concepts in each VP. So, we follow the following steps:

- **SAN**: represent the lexical similarity of the names of attributes. The DA (Distance of Attributes) measure the similarity between the attributes (respectively a_i and a_j) of two concepts (respectively C_1 and C_2).

$$\text{AD}(C_1.a_i, C_2.a_j) = \text{LSim}(C_1.a_i, C_2.a_j) \quad (8)$$

Every attribute of the concept C_1 is compared with every attribute of the concept C_2 :

$$\text{SAN} = 1 \text{ If two attributes are identical or synonymic, i.e.} \\ \text{AD} = 1 \\ = 0 \text{ the opposite case.}$$

- **SAD**: indicate the semantic equivalence of attributes $\text{Att}_j^{\text{VP1}}$ and $\text{Att}_j^{\text{VP2}}$ between the concepts of two VP according to the results of similarity calculated between the concepts in the previous stage and if C_j^{PV1} and C_j^{PV2} is already supposed similar (similarity of concepts) then $\text{SAD} = (\text{Min}(\text{Att}_j^{\text{PV1}}, \text{Att}_j^{\text{PV2}}) / \text{Max}(\text{Att}_j^{\text{PV1}}, \text{Att}_j^{\text{PV2}}))$ and $j \in [1, \text{Min}(C_j^{\text{PV1}}, C_j^{\text{PV2}})]$, in the opposite case $\text{SAD} = 0$.
- **SAV**: indicate the similarity between the values of atomic attributes, this factor = 1 if both values of both attributes are atomic and identical.
- **SFA**: this factor represents the average of the similarity between VP with the aim of their attributes; this factor is estimated when two concepts of two VP are considered similar by applying the measures defined in the previous phase. This factor is estimated by the following formula:

$$\text{SFA}(C^{\text{PV1}}, C^{\text{PV2}}) = \left[\sum_{j=1}^{\text{Min}(C^{\text{PV1}}, C^{\text{PV2}})} \text{SAD}(\text{Att}_j^{\text{PV1}}, \text{Att}_j^{\text{PV2}}) * [\text{SAN}(\text{Att}_j^{\text{PV1}}, \text{Att}_j^{\text{PV2}}) + \text{SAV}(\text{Att}_j^{\text{PV1}}, \text{Att}_j^{\text{PV2}})] / 2 \right] / \text{Max}(C^{\text{PV1}}, C^{\text{PV2}}) \quad (9)$$

- $\text{Min}(C^{\text{PV1}}, C^{\text{PV2}}$): the minimal number of attributes of every VP compared between the concepts ($C^{\text{PV1}}, C^{\text{PV2}}$).

- $\text{Max}(C^{\text{PV1}}, C^{\text{PV2}}$): the maximal number of attributes of every VP compared between the concepts ($C^{\text{PV1}}, C^{\text{PV2}}$).

Overview of classification algorithm: our proposed algorithm is based on the model of similarity by VP proposed above and includes the following stages: (O: the implied ontologies, B: the set of blocks after partitioning)

Algorithm: data sources classification

Input : O, B

Begin

For each ontology O_1 and O_2 from O//browse the set of

ontologies (*)

For $k=1..m$ do // browse the set of blocks

For each concept $c_1 \in O_1$ and $c_2 \in O_2$

Calculate Sim_G ; // calculate the global similarity between concepts

If $Sim_G = 1$ then

Create L ; // L represents a list of similar pairs of concepts

Add $\langle c_1, c_2 \rangle$ to L ; // according to the Sim_G result c_1 and c_2 are equivalent

For each attribute $att_x \in C_1$ and $att_x \in C_2$

Calculate SFA ; // calculate the global similarity between attributes

If $SFA = 1$ then $C_1 \equiv C_2$; // c_1 and c_2 are considered equivalent

Modify $MCVP$; // modify the compatibility matrix by VP

Calculate Sim_Mix ; // calculate the total similarity value

If $Sim_Mix \approx Average$ then // average: represent a defined value by the expert

Group O_1 and O_2 in the same category Cat_i

Delete O_2 from the classification process

Go to (*) and change the couple of ontologies to compare

End

Output: $Cat, L, MCVP$ (Cat : the set of categories after classification)

The algorithm has as a result a set of matrices of n couple of ontologies compared together called compatibility matrix by VP (Cf. fig 3)

		O_{i+1}				
Blocs		b_k^{PVj+1}	b_{k+1}^{PVj+1}		
O_i	b_k^{PVj}	$SimC$	$SimA$	$SimC$	$SimA$
	b_{k+1}^{PVj}					
					
					

Fig. 3 Compatibility matrix by VP (MCVP)

$$SimC = \sum_{i=1}^t Sim_G(C_1, C_2)$$

$$SimA = \sum_{x=1}^z SAF(C^{PV1}, C^{PV2})$$

- *The mixed similarity*: represents the last result of the similarity treatment, the value of this similarity is calculated from the compatibility matrix (MCVP). The mixed similarity consists in calculating the maximal number of concepts and visible attributes similar between every couple of ontologies (O_i, O_j). According to these results the categories of data sources are engendered.

$$Sim_mix = \sum_j \frac{[\text{Min}(C^{PV1}, C^{PV2}) + SimC + SimA]}{\text{Max}(C^{PV1}, C^{PV2})} \quad (10)$$

- $\text{Min}(C^{PV1}, C^{PV2})$: the minimal number of compared concepts between (VP_1, VP_2)

- $\text{Max}(C^{PV1}, C^{PV2})$: the maximal number of compared concepts between (VP_1, VP_2)

Example

Let we consider the example of automobiles society represented by a set of ontologies concerning mainly the concept « automobile » (Cf. Fig. 4). Every ontology includes the representation according to a particular viewpoint.

By applying the different steps of integration described in section 4 on the ontologies represented in Fig. 4, we achieve to group the ontologies $O1$ and $O3$ in the same category which represent the “Commercial VP” concerning the aspects of vehicle use (purchase, conduct, etc.). Also, the ontologies $O2$ and $O4$ are grouped in another category representing the “Physical VP” concerning external aspects of vehicles. Thus, as a result of this step we have two VP ontologies: physical VPO and commercial VPO.

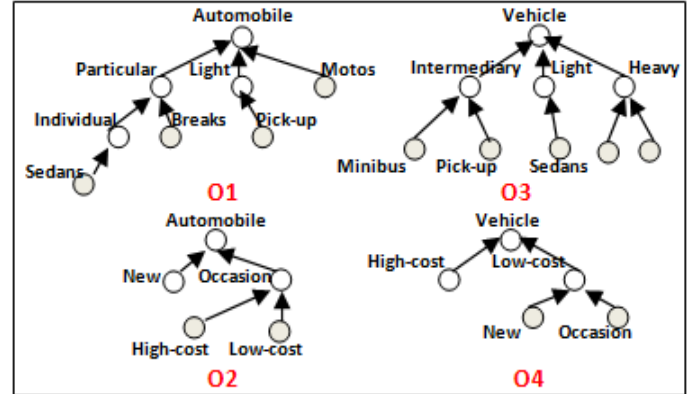


Fig. 4 Example of different representations of automobile society

B. Integration process

The second level of our approach concerns the realization of two successive types of integration:

- 1) *Pre-Integration*: The data sources of every category are integrated into a local ontology, called later “VP Ontology”. The VP ontologies are built with the GAV approach and provide a unified view of the data sources grouped in each category. Such integration is reused for every category of data sources.
- 2) *Global Integration*: this level is represented by the domain ontology or global ontology (GO), which reconciles all the aspects existing in this domain. The GO plays the role of a global scheme and every VP ontology can be seen as a specification of this ontology.

V. THE PROPOSED ARCHITECTURE

The architecture that we propose which rises from the suggested approach is built on three levels presented in Fig. 5, every level is articulated on its own components described in the following:

A. Data Sources Level

This level represents the data sources which contain their own ontologies following different viewpoints. Because of their same data structure, the structural heterogeneity is no more difficult for data integration. The data sources follow different interpretations of objects where the same concept can be defined according to different viewpoints.

B. Viewpoint Integration Level

This level is essentially constituted of:

VP ontology: The VP ontology gathers all the concepts concerning a particular VP, represented in a distributed way in

the data sources. It is conceived according to the GAV approach.

VP Mediator: plays intermediary's role between the global mediator (presented below) and the VP ontologies. It allows facilitating the localization of the adequate data sources to answer a request.

Local mapping table (TML): It allows representing the possible links between the local ontologies and the VP ontologies. At this level, the Global as View approach is used in our mediation system to describe the correspondences between schemes (data sources schemes / VP ontologies).

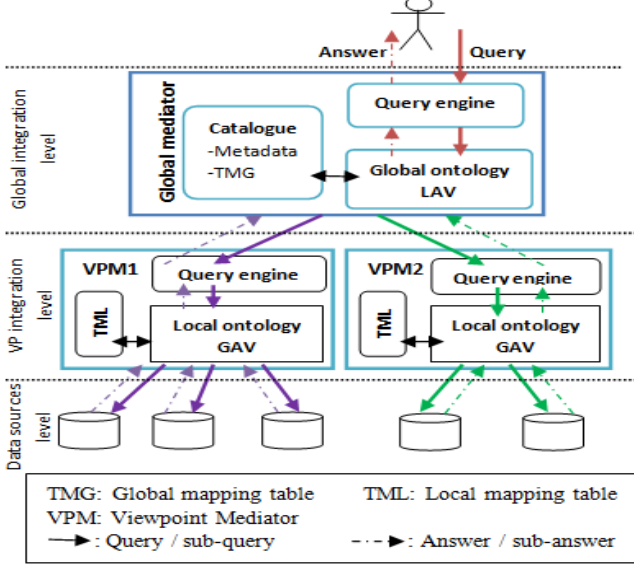


Fig. 5 The mediation architecture

C. Global Integration Level

Represents the final level of integration based mainly on three essential components:

Global mediator: it gives an integrated vision of the various VP ontologies schemes. The user expresses his requests in the terms of the global ontology. These requests are successively refined by the global mediator that specifies what the aimed VP ontologies are, and then it belongs to the VP mediators to identify the adequate data sources to be questioned.

Global ontology: represents the vocabulary base for the users taking into account the various perceptions of the actors implied in the domain. The articulation at this level indicates relations between the concepts of the global ontology and those of the VP ontologies, which are defined according to LAV approach.

Catalogue: It contains global mapping table and metadata.

- *Global mapping table (TMG)*: it allows to describe the various correspondences between the global ontology (GO) and the VP ontologies (OVP). It permits to represent for each concept in the global ontology the corresponding concepts in every VP ontology.
- *Metadata*: describes the capacities of the VP ontologies / data sources, and information on their format. Other information on the viewpoints of the data can be stored at this level. These metadata helps well the requests engine in

the optimization of the requests plans and limits the research only in the relevant sources.

VI. QUERY TREATMENT

A user request is expressed in terms of the GO, we obtain easily a reformulation in sub requests in terms of the OVP because this activity will be reduced and concerns only the VPs implied by the user's request. The answer to these sub requests is easier to realize because their decomposition concerns a reduced number of data sources. Thus, one of the advantages of the classification by VP is the reduction of the number of data sources questioned.

Definition of a request execution plan: (PRE)

In our approach, the activity of query rewriting will be made on two levels:

- At the global mediator level, this uses metadata associated to determine the corresponding VPO.
- At the VP mediators level, where every mediator concerned by a sub request decomposes it on the local data sources implied by this sub request.

A requests execution plan (PRE) represents all the rewritings found by the mediator, and which are equivalent to the user's request.

Example:

Let use some terminologies from the example represented in Fig.4. With $V = \{V_0, V_1, V_2, V_3, V_4, V_5, V_6, V_7\}$ as a set of views with the following definitions:

- def (V_0) = Automobile
- def (V_1) = AssociatedSituation
- def (V_2) = Automobile \cap Particular
- def (V_3) = Automobile \cap Occasion
- def (V_4) = Automobile \cap New
- def (V_5) = Particular \cap Sedans
- def (V_6) = Particular \cap Breaks
- def (V_7) = Automobile \cap Intermediary

Let be the following request:

$Q(Y) : \text{Automobile}(Y) \wedge \text{AssociatedSituation}(Y, L) \wedge (\text{Automobile} \cap \text{Particular})(L) \wedge (\text{Automobile} \cap \text{New})(L)$. It expresses the search for automobile, such as the automobile is new and particular. We can deduct that the user's search combines two characteristics of two different viewpoints:

New \leftarrow From the commercial VP.
Particular \leftarrow From the technical VP.

Let be the following assertions, described from the inter-VP rules:

- R1 : $\forall \text{Particular}(X), \exists \text{Intermediary}(Y) \mid X \perp Y$
Particular \cup Intermediary
- R2 : $\forall \text{New}(X), \exists \text{Occasion}(Y) \mid X \perp Y$
New \cup Occasion

By using these two rules, we can eliminate some definitions of the given views, to define the execution plan for the request Q. The set of eliminated views is represented by $V_E = \{V_7, V_3\}$.

The following scheme represents an example of query execution (q_v is the expansion of the request using the definition of views described above):

$Q(Y) : \text{Automobile}(Y) \wedge \text{AssociatedSituation}(Y, L)$
 $\wedge (\text{Automobile} \sqcap \text{Particular})(L) \wedge (\text{Automobile} \sqcap$
 $\text{New})(L)$

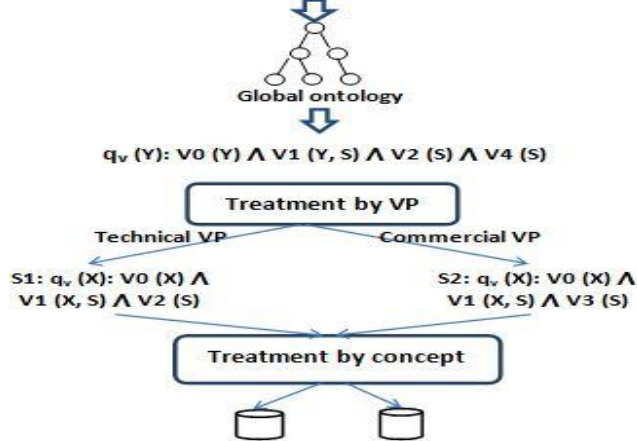


Fig. 6 Request treatment by VP

VII. COMPARAISON STUDY

The most of the integration systems adopt the architecture containing mediators and adapters as defined by [16]. The originality of our approach is that in addition of the characteristics of the current systems, seeks to promote an architecture which improves the integration of heterogeneous data sources by supporting the notion of VP.

Our approach is characterized by architecture in two coats of mediators. The inferior coat is established by mediators VP that use the GAV approach to integrate the schemes of a data sources category representing the same VP. The superior coat has a single global mediator who integrates schemes specialized the inferior coat using the LAV approach. Thus, our approach benefits from the following characteristics:

- The conception of our system reflects the need to create a system which is strongly independent from the number of sources.
- The problem of a new data source addition implied by the use of the GAV approach will be avoided in our approach where: a new data source will be taken into account by a specific category of sources, in other words this source is built like a view on the corresponding VP ontology without affecting the global ontology.
- It is known that the complexity of the global view increases with the number of data sources and that the GAV approach is more effective and more practicable if the number of data sources to be integrated is small and stable. These two qualities are realized in our approach. Indeed, the VP mediator will always have to integrate a reduced number of partial schemas.
- The requests are expressed in terms of the global scheme and their execution is made by reaching every source concerned by the VP mediator. According to [14], the price to be paid for the flexibility and the simplicity of the mediator updates augment with the increase of the sources number. In our architecture, this complexity will be reduced because every VP mediator has a compact number of data sources to be integrated. So, the rewriting of requests in terms of views will be easier.

VIII. CONCLUSIONS

In this paper we have proposed a viewpoint-based approach of integration which introduces the combination of two successive approaches of integration: GAV and LAV approaches. In a future work, we plan to develop an algorithm based on the characteristics of viewpoints; essentially the notion of "inter-VP rules" to improve the performances of queries treatment.

REFERENCES

- [1] F. Benchikha, M. Boufaïda, "The Viewpoint Mechanism for Object-Oriented Databases Modeling Distribution and Evolution". *Journal of Computing and Information Technology*, CIT, Vol.15, n°2, Sep. 2007 pp 95-110.
- [2] D. Beneventano, S. Bergamaschi, S. Castano, A. Corni, R. G. Guidetti, G. Malevezzi, M. Melchiori, and M. Vincini, "Information integration: The MOMIS project demonstration", *International Conference on Very large Data Bases*, Cairo.Egypt, VLDB, Sep. 2000, pp. 611-614.
- [3] A. Boussis, "Intégration de sources de données à base ontologique dans un environnement P2P", M. S. thesis, Graduate National School of Computing E.S.I, 2008.
- [4] A. Bouzouan, B. Bouchard, "Human-Agent Viewpoint Similarity based on Terminological Logic for Mixed-Initiative Planning", *Intelligent Agent Technology*, Halifax.Canada, IAT, Oct. 2003, pp. 482-485.
- [5] S. Buss, R-D. Kutsche, U. Leser, and H. Weber, "Federated information systems: Concepts, Terminology and Architectures", Technical Univ., TU Berlin, Tech. Rep., NR 99-9, App.1999.
- [6] X. Dung Nguyen, "Intégration de bases de données hétérogènes par articulation a priori d'ontologies : application aux catalogues de composants industriels"; Ph.D dissertation, Poitiers Univ., France, 2006.
- [7] E. Guérin, "Intégration de données pour l'analyse de transcriptome : mise en oeuvre par l'entrepôt GEDAW (Gene Expression Data Warehouse)", Ph.D dissertation, Rennes 1 Univ., S.V.E, France, 2005.
- [8] W. H. Inmon, "Building the Data Warehouse". 3rd ed, New York, 1996.
- [9] Z. Laomo, "An extended hybrid ontology approach to data integration". *International Conference on Biomedical Engineering and Informatics*, Tianjin.China, BMEI, Oct. 2009, pp 1-4.
- [10] M. Lenzerini, "Data integration: a theoretical perspective", *Symp. Principles of Database Systems*, Wisconsin. USA. PODS, Jun. 2002, pp 233-246.
- [11] D. Lin, "An Information-Theoretic Definition of Similarity". *International Conference on Machine Learning*, Wisconsin. USA, ICML, Jul. 1998, pp 296-304.
- [12] E. Mena, V. Kashyap, A. Sheth, and A. Illarramendi, "OBSERVER: An approach for query processing in global information systems based on interoperation across pre-existing ontologies". *International Conference on Cooperative Information Systems*, Brussels. Belgium, CoopIS, Jun. 1996, pp 14-25.
- [13] P. Resnik, "Using Information Content to Evaluate Semantic Similarity in Taxonomy", *International Joint Conference on Artificial Intelligence*, Québec. Canada, IJCAI, Aug. 1995, pp. 448-453.
- [14] M.-Ch. Rousset, A. Bidault, Ch. Froidevaux, H. Gagliardi, F. Goasdoue, Ch. Reynaud and B. Safar, "Construction de médiateurs pour intégrer des sources d'information multiples et hétérogènes : le projet PICSEL". *Journal of Information - Interaction - Intelligence*. I3, Vol.2, n°1, Jan. 2002, pp 5-59.
- [15] J.J. Shilling, P.F. Sweeney, "Three Steps to Views: Extending the Object-Oriented Paradigm". *Conference on Object-Oriented Programming: Systems, Languages, and Applications*, New Orleans. Louisiana, OOPSLA, Oct. 1989, pp. 353-361.
- [16] G. Wiederhold, "Mediators in the architecture of future information systems", *Journal of IEEE Computers*, Vol.25, n°3, Mar. 1992, pp 38-49.
- [17] WordNet (2012, December 27). Princeton Univ. [Online]. Available: <http://wordnet.princeton.edu/>