

Managing Difference of Opinion in Semantic Structures

Samuel Gesche

LIRIS

INSA de Lyon

69621 Villeurbanne Cedex – France
00 33 - 04 72 43 63 48

Samuel.gesche@insa-lyon.fr

Guy Caplat

Département Informatique

INSA de Lyon

69621 Villeurbanne Cedex – France
00 33 – 04 72 43 82 82

Guy.caplat@insa-lyon.fr

Sylvie Calabretto

LIRIS

INSA de Lyon

69621 Villeurbanne Cedex – France
00 33 – 04 72 43 88 94

Sylvie.calabretto@insa-lyon.fr

ABSTRACT

Evaluation of automatic indexing is a common practice. Whenever new algorithms are created to generate semantic structures, it is necessary for their designers to demonstrate how better they are in the particular context for which they have been developed. However, the evaluation depends on the availability of a result of reference. Should such a reference be missing, evaluation is replaced by confrontation. There the structures themselves have to be compared instead of shared metrics, sometimes without even the option to choose the best one. Our work takes place in this context. We present a method for human-driven confrontation of specialized semantic structures. Our aim is to improve the user's ability to use another person's work to enhance his.

Categories and Subject Descriptors

H.1.2 [Information Systems]: User/Machine Systems – *Human Information Processing*

General Terms

Theory, Human Factors, Verification

Keywords

Confrontation, Viewpoints, Human-Driven Evaluation

1. INTRODUCTION

Structuring knowledge, whether by simply listing resources or by designing huge formal ERPs, is one of the tasks that human beings are the less reticent to give away to computers. It is, indeed, often considered as a step that is as boring as unavoidable, and according to the stereotype it will eventually be trainee's work to complete it.

Assigning a task to a computer is usually done using at some level two orthogonal approaches. In the first one, the computer will, in a batch process, perform the task without the help of any human except for entering the initial data. In the second one, an

interactive process will let the human perform himself the task. It will provide an extended computing power to make instantly small, repetitive subtasks. In the process of indexing, the first approach is assigned for example to the huge Internet indexing robots. The second one is used for manual organization of files into a file system.

Whatever approach is chosen, the post-process of evaluation plays an important role in the design of such software. New algorithms have to be proven better, even in a restricted field, than older ones, and generally speaking the computer must either perform the task faster, or better than a human would.

This implies, in the batch –and therefore automatic– approach, to transfer to the computer the knowledge of a human being. Indexing is inseparable of search and retrieval, for which it is designed. So indexing systems are coupled with information retrieval systems, and they have to answer as well as possible to well-prepared examples, which means as close as possible to a human-made reference –the only available reference. And worldwide conferences are held where every system can be compared with the others (for instance [10]).

In the second approach, which is interactive, the concern is different. It is not the ability to build that is important, but the ability to help building. Evaluation of the system does not play such an important role –with the exception of user-friendliness. But another issue is raised: when a human –or a system– needs ten years to build an index of a given corpus, it is de facto acknowledged a reference. When every user is able to perform the same task within a working day, it is merely considered as a viewpoint.

This is for instance what appears in the world of ontologies, especially in the Semantic Web [5]: the same problem can be solved by different communities or companies around the world, which build each their own ontology. These ontologies may be only slightly different, but in a formal context this is sufficient to raise the issue of heterogeneity. The domain of ontology alignment and matching is dedicated to solving the first part of the problem: managing heterogeneity in semantic structures designed for automated processing. Our work addresses the other part: managing heterogeneity in semantic structures designed for human use [3].

We will start by presenting some approaches related to our concern. Then, we will describe our contribution, the theoretical context of the management of difference of opinion in semantic structures designed for human uses. We will also shortly present

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SADPI'07, May 21-22, 2007, Montpellier, France.

Copyright 2007 ACM ISBN 978-1-15159-668-4.

what has been currently implemented. Finally, we will give some perspectives about our future work.

2. RELATED DOMAINS

2.1 Information Retrieval Evaluation

Information Retrieval is a domain where indexing and structuring documents play a great role. Most of the solutions that involve computers are all-automated. Documents and queries are given as input, and relevant document are returned as output. One of the concerns is thus to evaluate the algorithms that transform input into output: processes of indexing, query processing and eventually information retrieval.

While evaluation itself remains an open problem, there are several methods that are currently used. Most of them impose the use of pre-answered queries. Thus, given the amount of necessary data, such methods give place, every year, to evaluation rounds dedicated to some type of data, such as [6] and [12] for multilingual search, [10] for search in XML and TREC [16] which is itself subdivided according to several types of data (video, legal, genomics...).

The common requirement for these campaigns is to build a test environment, where performances of all participants can be compared. Since the best system is yet to be discovered, the human judgment prevails when it comes to choosing the good and bad answers to the test queries.

Information Retrieval Evaluation is close to our concern, except that this technology is dedicated to structures produced by computers, as shown in figure 1.

Producer \ User	Human	Computer
Human	Viewpoints Confrontation	Ontology Alignment
Computer	IR Evaluation	IR Evaluation

Figure 1. Structures Evaluation Technologies.

2.2 Ontology Alignment

Ontologies are at the heart of the Semantic Web [5]. They are “explicit specifications of conceptualizations” [9], and they have proven to be able to provide a formal context to semantics, in order to use digital processing at a level where other means (like simple word indexing) fail. Besides, ontologies are, for those who are used to modeling, a way to translate their knowledge in a format which is easily shared. Therefore, when building an ontology, the stress is often put on the reusability and the need of a consensus [15]. In other words, an ontology is not an opinion relative to a person, but the description of a part of the real world, which has to be as indisputable as possible.

That is however only the theory that underlies the concept of ontologies. The reality is often different. Whenever several communities work in the same area, they build as much different ontologies as there are communities. The concept of ontology alignment has emerged to solve this issue [7]. Its aim is to find connections between them and to generate a set of mapping

relations. Ontology alignment thus allows interoperability between ontology-based services or applications [17].

Ontology alignment also comes close to our problem. Sometimes it is even related to multi-viewpoint approaches [18]. However, we work on structures designed for human use, whereas ontology matching is designed for structures that will underlie a computer process. Some techniques for ontology matching may therefore be inadequate.

3. CONTEXT

3.1 Presentation

We work on semantic structures made by and for human users. The digital process we consider is fully dedicated to the enhancement of human abilities. The ARTCADHi team [14] already designed an environment, Porphyry, for the expression of expert viewpoints. We call them opinion-viewpoints, for reasons that we will expose later in this section. This environment allows researchers in the area of humanities to work in a digital context without the need of a computer specialist.

3.2 Description Networks

Since the users of Porphyry are not acquainted with the paradigm of modeling, the language in which they express their viewpoints is made as simple as possible, and at the same time as expressive as possible. The counterpart is that the formalism is very poor. It is a graph-based formalism, involving a single relation type, specialization (with all the fuzziness that a researcher can put in that), and a node type, descriptor, subdivided according to the ability of its instances to be generalized and/or specialized. The whole metamodel is presented in figure 2.

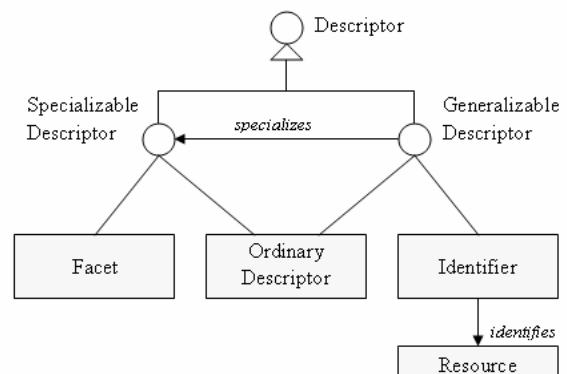


Figure 2. Metamodel of the Descriptions Networks.

This language is currently evolving towards a richer formalism, but the need remains for a formalism which can be understood and readily used by experts that are not familiar with the paradigm of modeling.

3.3 Opinion-Viewpoints

For the purpose of confrontation, we defined the notion of opinion-viewpoint as opposed to the notion of viewpoint which

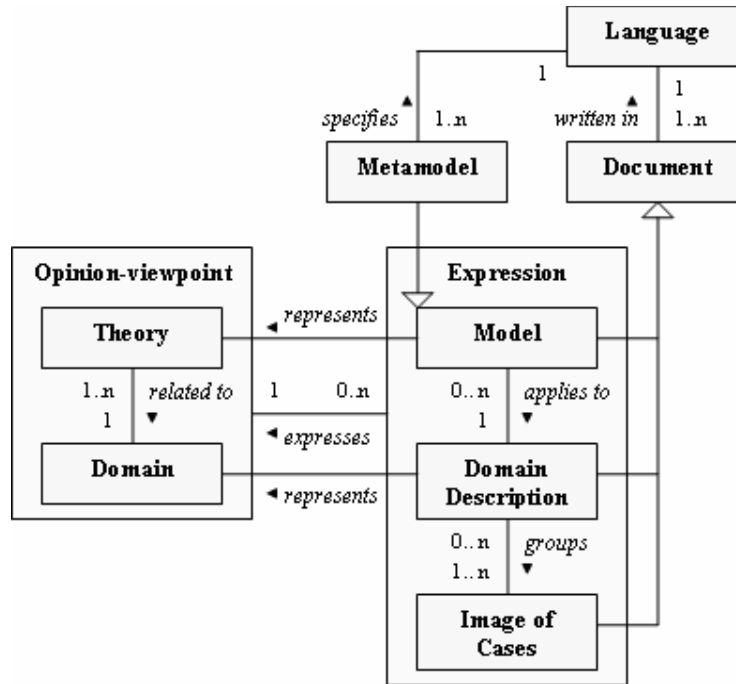


Figure 3. Model of the Opinion-Viewpoint and its Expression.

is an emerging paradigm in Computer Science and especially in Information Systems Design (as a static, user-oriented perspective on an entity: [8], [11]). The term itself is taken according to [13]. An opinion-viewpoint is a dynamic, non-consensual theory which is expressed on a domain for the purpose of sharing. It can be easily found in Sciences in the stage of theory elaboration, and, to give an immediate example, this subsection is the expression of an opinion-viewpoint on viewpoints.

Such a theory usually has a part which is consensual, and another part which depends on the author, his researches, his intuition and his background. The personal part is most important when the domain covered by the theory is emerging, or when there are few experts working on it.

Opinion-viewpoints are expressed by models, which are written in a defined language. Generally speaking, the language can be anything suitable to underlie communication, but here we limit ourselves to the formalisms used to build knowledge structures, and they are often graph-based. Figure 3 describes our model for the opinion-viewpoint and its expression.

The results of an indexing process (indexes or more semantic structures such as thesauri or ontologies) do not aim at being opinion-viewpoints. They have to express consensual knowledge, since the important viewpoint when accessing document is the viewpoint of the user and not the indexer's. However, creating and filling such structures with the aim of structuring knowledge is highly dependent on the person or the algorithm that leads to it. Whenever there are different structures for the same purpose, they are likely to be different and so they can be considered as opinion-viewpoints. And generally speaking, we note that as soon as knowing who made a structure is important, this structure is no longer out of the scope of the concerns about opinion-viewpoints.

3.4 Objective

The objective of the ARTCADHi project is to create an environment, Porphyry, dedicated to researchers in humanities. The project succeeds if they can use the environment to enhance their job [2]. The machine allows them to do faster and better the work that it is built for: repetitive, formal tasks. The environment will let them do by themselves the tasks that require expert knowledge, and sometimes even define the use of the functionalities [2]. And like a word processor makes the writing of letters more efficient, this shall make their research more efficient.

Porphyry is mainly used by teams studying archaeological corpuses. Therefore, the environment provides tools oriented towards multimedia corpuses structuring. Building such structures, using the formalism of Description Networks, is already a functionality of Porphyry.

4. CONFRONTATION

4.1 Human-Driven Confrontation

The functionality of opinion-viewpoints confrontation meets a need in any researcher's job. In our case, it is thus a human-driven process.

The choice of a human-driven process versus a machine-driven process is often made when the state of the art does not permit to make the process automatic, or when the process itself is not formalized enough to allow the use of black box algorithms. In our case, both are true: even if the research is made according to a method that is formalized enough to be taught, it requires a scientist to use this method. And in these areas of knowledge where definitions themselves are not so definitive, and where no

thesaurus can even be built [4], an automatic process is irrelevant. This is summarized in figure 4.

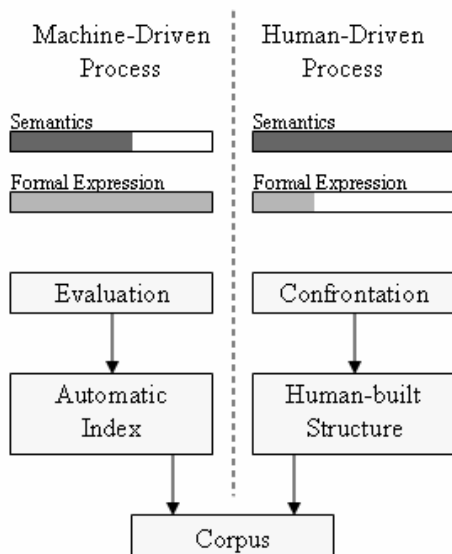


Figure 4. Machine vs Human-Driven Process.

Besides, since the researchers who shall use our system are not used to the paradigm of modeling, they tend to be less careful in putting the semantics in their documents using the right primitive. That would lead any automatic process to make mistakes that can be more easily avoided by putting one of these researchers at the commands than by trying to model their activity.

The counterpart of all this is that we must build an environment for these researchers, with notably an adapted GUI. Besides, we have to be careful not to leave too time-consuming tasks block the researchers' work. If a five minutes process is an honorable performance for an ontology alignment, it can be too much time in our system. Especially since several iterations may be necessary for the expert to set all parameters correctly.

4.2 Principle

The overall purpose of confrontation is to make new ideas emerge, by allowing a researcher to see the differences between his and his fellow's opinion-viewpoint.

We thus have to find where the models show differences of opinion between their authors. There are two main kinds of such differences:

- Either some of the knowledge is expressed in one of the models and not in the other: in this case, it is interesting to know whether models cover the same domain, and to which point the lacking knowledge could enrich –and maybe modify– the model.
- Or, part of the knowledge is expressed differently in the models: in this case both authors can enrich their model if they understand the other's approach.

Some other concerns can also be useful while comparing a researcher's viewpoint with another's. However, they are best

left to the expert's judgment, unless it is specified in the opinion-viewpoint itself:

- Cultural gap: are the authors close culturally speaking?
- Temporal gap: are their opinion-viewpoints contemporary?
- Subject gap: are their domains of research or the subjects they study close?

4.3 Model

The process of confrontation of opinions is constituted by a succession of actions. To describe all the confrontations requires defining a language of confrontation. It is a language to express these actions and the notions on which these actions act. We can then define a model of confrontation for the confrontation of a set of models.

The language of confrontation shall be independent of the language in which the models are written, or adapted to all languages used.

To confront two models, we need, as shown in figure 5:

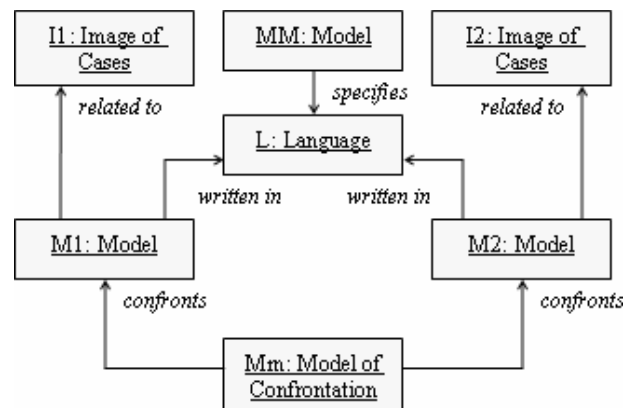


Figure 5. Model of Confrontation.

- Two models $M0$ and $M1$ to confront, written in the same language L , specified by the model MM . They are related to the images of cases $I1$ and $I2$ (images of cases can be considered as sets of expressions of instances that are outside of the model, i.e. external resources);
- A model of confrontation Mm specifying all the actions of confrontation. This model is the model of the process and not its result. Applying the model gives the result.

This model Mm has to be included in the software only as far as the process of confrontation is automatic. In our case (see next section), this model underlies the organization of the GUI and does not appear elsewhere in the environment, because confrontation is made by the user. However, the need remains of a metamodel for the language L in which the models are expressed, because these models have to be presented to the user in a form allowing confrontation in the environment.

4.4 Multi-language and multilingual confrontation

Opinion-viewpoint confrontation cannot be limited to a single language. Even if our work is made in the context of communities using the Description Networks and the Porphyry

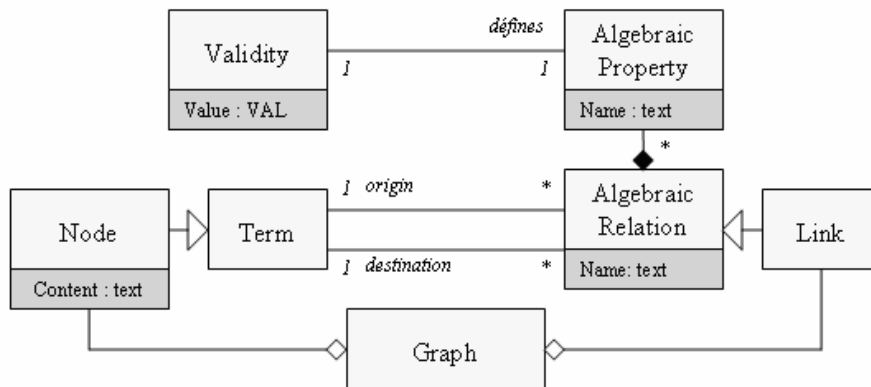


Figure 6. Model of a formalism satisfying our hypotheses.

environment, we are aware of other formalisms created to express theories in the humanities, like SDC [1]. And confrontation of opinion-viewpoints is not a problem that only occurs in the area of humanities.

Our model only works with single-language confrontation, so opinion-viewpoints models have to be translated into a single formalism that can be used by the process of confrontation. We briefly present in the next section a unified language for confrontation.

The good news is that multilingual confrontation is less a problem than multi-language confrontation, provided that the user is able to work with both languages. Since it is the user that actually confronts the opinion-viewpoints, the language in which it is expressed (provided that the formalism is the same) is not a limit for the computer (even if part of the functionalities may not function properly).

5. IMPLEMENTATION

5.1 Leucippus

Opinion-viewpoints are expressed in various languages that are translated into a specific language. Some hypotheses need to be respected:

- All these languages of expression are graph formalisms. All other formalisms are banned from our field of competence. Besides, we need for our approach to be able to get a metamodel or a specification of each of the formalisms. The opinion-viewpoints are thus expressed in the form of models written in graph formalisms.
- We consider that the graphs only convey text content (or URIs) in the nodes and links.
- We plan to base part of our analysis on the algebraic properties of the links (this has not been done at this point because we mainly work on the Description Networks in which there is only one relation type). Therefore, we need to know these properties.

Finally, a graph satisfying our hypotheses has a metamodel such as the one presented in figure 6.

Using these hypotheses, we have created a unified language, called Leucippus, for translating opinion-viewpoints in order to confront them. Other languages, if they satisfy the hypotheses, can be translated into Leucippus. This ensures that the data that we need for confrontation are saved, even if a part of the semantics is lost (in any translation there is a loss of semantics).

Languages that do not satisfy the hypotheses can also be translated, but the less they are conforming to them, the more semantics will be lost. Structures written in a language satisfying our hypotheses can be translated into Leucippus using graph transformation. There is a XML file format for Leucippus, so this transformation can be made outside of the environment, as a plug-in.

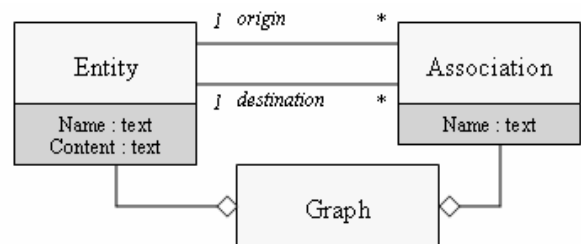


Figure 7. Metamodel of Leucippus.

Leucippus has the metamodel as presented in figure 7: a graph written in this formalism contains a set of entities and a set of associations. An association is an oriented relation: it has an entity as origin and another one as destination.

As shown in figure 8, there are six types of entities and three types of associations in Leucippus. An example of a graph translated into Leucippus is given in figure 9.

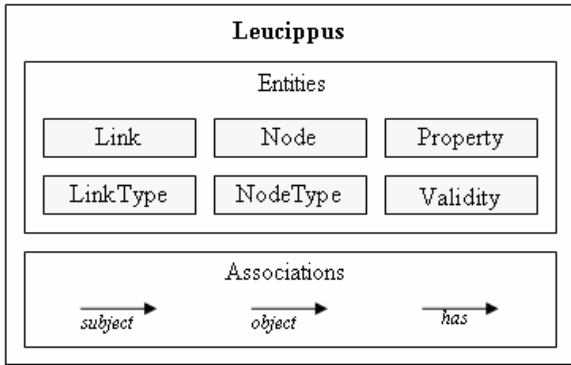


Figure 8. The notions of Leucippus.

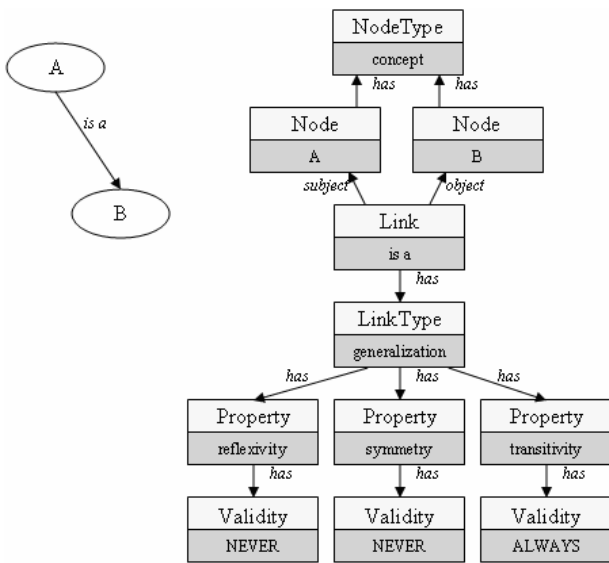


Figure 9. A graph projected into Leucippus.

5.2 Platon

Based on the Leucippus Language that provides a unified language for confrontation, we designed a prototype for the purpose of testing our approach. We call it Platon.

Platon allows the user to work on a set of opinion-viewpoints expressions (graphs) and to prepare them for confrontation by disambiguation. Eventually, confrontation is made and a result is given as a table.

Graphs disambiguation allows keeping the confrontation process simple. Since currently relation type semantics is not taken into account, the confrontation algorithm is as follows:

- Search for triplets (node/link/node) that are common to at least two graphs;
- For all these triplets, search in all graphs the triplets (node/link/node) where at least one node is common.

The principle is to search for interesting parts of the graph (by finding common triplets) and then to show which graphs have something related to these parts but maybe different.

Graphs disambiguation is necessary for defining when nodes are the same, and when they are different. Since our users are not specialists of linguistics, we use fuzzy definitions for the concepts of synonymy and homonymy (for these definitions, we call “terms” the content of nodes, which can actually be more than a single word, and “words” the sequence of letters).

- Two terms are said “synonyms” whenever they are exchangeable. It can be true synonymy or they can mean the same only in the context of confrontation.
- A term is said ‘homonymic’ whenever two instances of this term (in the particular context of the confrontation) could have distinct meanings. This includes actual homonymy, polysemy and all-purpose words such as “miscellaneous” or “others”.

It is hard for an algorithm to predict “homonymy”, so the choice is left to the user alone, only focusing him on duplicate terms. For the “synonymy”, however, there are more possibilities, so we put a (non-optimized at this time) algorithm for the purpose of suggesting synonyms. This algorithm studies the terms at the level of the character sequence. To give better guesses, the user is brought to find the words that give no meaning to the terms (for example, if *in the hall* and *hall* refer to the same thing, it is necessary to define that *in* and *the* are in this case meaningless).

Figure 10 presents all the functionalities that are offered to the user. Figure 11 shows an example of confrontation result (we confronted the geographic division of Greece for archaeological excavations results reports along several decades).

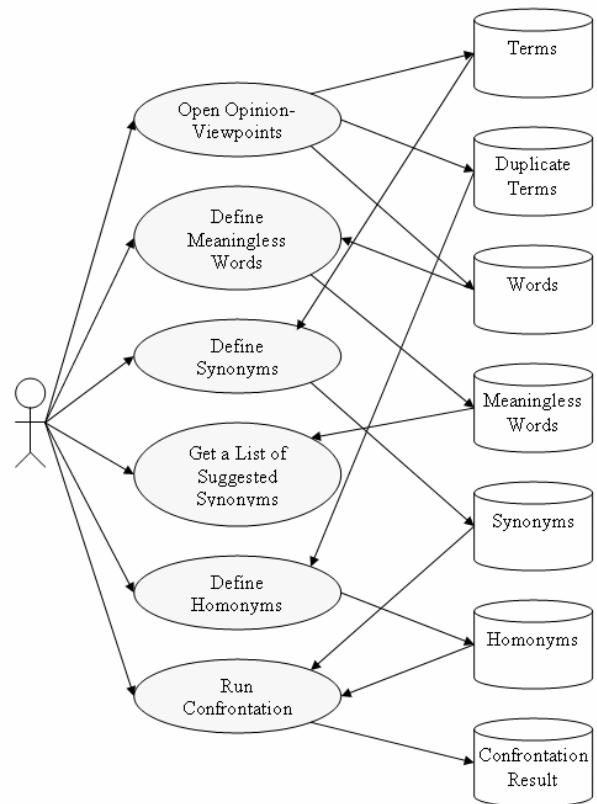


Figure 10. Use Case Diagram of Platon.

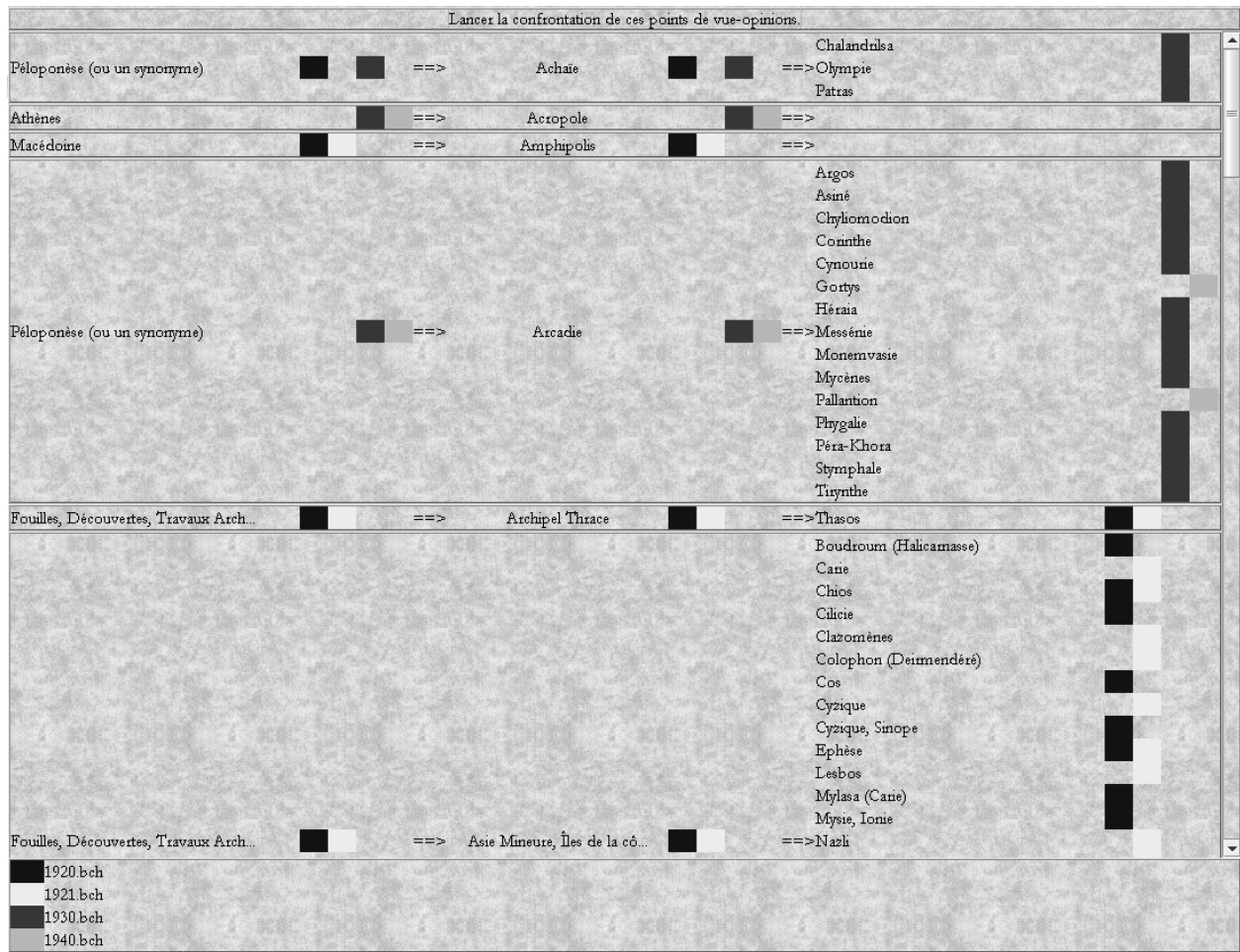


Figure 11. An example of Confrontation Result.

6. CONCLUSION

We have presented an approach for the management of difference of opinion in semantic structures by confrontation. The design of the structures is human-driven as well as the confrontation itself, and the aim is to enhance human abilities rather than make the machine as intelligent as the human.

The current state of our research allows us to perform the process of confrontation among opinion-viewpoints written in simple graph-based formalisms such as Description Networks. Our concern is now the further study of algorithms used in ontology alignment, because some of them can be used in opinion-viewpoint confrontation as well.

We work now on the extension of the prototype to more complex languages, taking into account the advances in ontology alignment.

7. REFERENCES

- [1] Aussenac-Gilles, N. Ontology or meta-model for retrieving scientific reasoning in documents: The Arkeotek project. *In Workshop on exploring the limits of global models for*

integration and use of historical and scientific information (2006).

- [2] Bénel, A. Porphyry au pays des paestans : usages d'un outil d'analyse qualitative de documents par des étudiantes de maîtrise en iconographie grecque. *In Actes du colloque international "Corpus en Lettres et Sciences sociales : des documents numériques à l'interprétation", Albi, juillet 2006.* Texto!, june 2006, vol. XI, n°2. p.182-189
- [3] Bénel, A., Calabretto, S., Iacovella, A. and Pinon, J.-M. Porphyry 2001: Semantics for scholarly publications retrieval, *In Proceedings of the thirteenth International Symposium on Methodologies for Intelligent Systems, Lecture Notes in Artificial Intelligence (2002)*, Springer-Verlag, 2366, pp. 351-361.
- [4] Bénel, A., Eyged-Zsigmond, E., Prié, Y., Calabretto, S., Mille, A., Iacovella, A., Pinon, J.-M. Truth in the Digital Library: From Ontological to Hermeneutical Systems. *In: Proceedings of the fifth European Conference on Research and Advanced Technology for Digital Libraries, Darmstadt, September 4-9, 2001.* Lecture Notes in Computer Science #2163. Berlin: Springer-Verlag. p.366-377.

- [5] Berners-Lee, T., Hendler, J. and Lassila, O. The Semantic Web. *Scientific American*, 2001.
- [6] Cross-Language Evaluation Forum (CLEF), <http://www.clef-campaign.org>
- [7] Ehrig, M. and Staab, S. QOM - Quick Ontology Mapping. *The semantic web*, Lecture notes in computer Sciences (2004), Springer Verlag, pp. 683-697.
- [8] El Asri, B., Nassar, M., Coulette, B., Kriouile, A. Assemblage de composants multivues par contrats. In *actes du XVIIIème congrès INFORSID (2005)*, pp 29-44.
- [9] Gruber, T.R. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, V (1993), 2, 199-220.
- [10] INitiative for the Evaluation of XML Retrieval (INEX project), <http://inex.is.informatik.uni-duisburg.de>.
- [11] Lahna, B., Khayati, O. Towards a viewpoint-based design approach. In *Proceedings of CSITeA'04*.
- [12] NII Test Collection for IR Systems (NTCIR), <http://research.nii.ac.jp/ntcir>.
- [13] Ribièrè M., Dieng R. A Viewpoint Model for Cooperative Building of an Ontology. In *Proceedings of the tenth International Conference in Conceptual Structures*, Lecture Notes in Computer Science (2002), Springer-Verlag, 2393, pp. 220-234.
- [14] The Porphyry Project, <http://www.porphry.org>.
- [15] Sampson, J., Lanzenberger, M. Visual Ontology Alignment for Semantic Web Applications. *Advances in Conceptual Modeling - Theory and Practice*, Lecture Notes in Computer Science (2006), Springer-Verlag, 4231, pp. 405-414.
- [16] Text Retrieval Conference (TREC), <http://trec.nist.gov>.
- [17] Zghal, S., Kamoun, K., Ben Yahia, S., Mephu Nguifo, E. and Slimani, Y. EDOLA : Une nouvelle méthode d'alignement d'ontologies OWL-Lite. In *Proceedings of the Conférence en Recherche d'Information et Applications (CORIA'07)*, pp. 351-366
- [18] Zhang, K., Hu, Y., Wang, Y. Multiple Viewpoints Based Ontology Integration. In *Proceedings of Grid and Cooperative Computing: Second International Workshop, GCC 2003, Shanghai, China*. Lecture Notes in Computer Science, Volume 3033 / 2004, Springer Verlag, pp. 690 – 693