

# Alignment Methodological Approach of Evolving Domain Sub-Ontologies using Terminological and Structural Matchers Applied to Tourism Domain

Khalid Tatane  
GMES Laboratory,  
National School of  
Applied Science  
B.P: 1136/S Agadir-  
Morocco

Brahim Er-Raha  
GMES Laboratory,  
National School of  
Applied Science  
B.P: 1136/S Agadir-  
Morocco

Chihab-eddine  
Cherkaoui  
Laboratory IRF-  
SIC, Faculty of  
Science Agadir  
Agadir, B.P: 37/S  
Agadir-Morocco

Sanaa Mouhim  
Laboratory IRF-  
SIC, Faculty of  
Science Agadir  
Agadir, B.P: 37/S  
Agadir-Morocco

## ABSTRACT

The use of ontologies to resolve problems of semantic interoperability has become a considerable challenge. To pool distributed and shared knowledge, the handling of ontologies requires defining “*set of matches*” to align them. This alignment process of ontologies is deemed to be complex as it is based on the measure analysis of similarity (Matchers). We suggest in this paper the classification of different alignment techniques combining at the same time several methods to generate semi-automatically Mapping. Opposite to these techniques, we present helping methodological approach of sub-ontologies alignment. We propose a realistic approach adapted to tourism domain which covers many sub-domains.

## Keywords

Global ontology, domain sub-ontologies, ontologies alignment, Similarity measuring, Mapping, Matcher, Semantic interoperability.

## 1. INTRODUCTION

This research is carried out in IRF-SIC<sup>1</sup> Laboratory, especially the SIC<sup>2</sup> team in the framework of GECO-WES<sup>3</sup> project which aims the development and use of semantic Web technologies based on: ontologies, collaborative work and E-learning to manage knowledge in organizations. First researches in this project allowed the setting of first ontological kernel for tourists (S.Mouhim et C.cherkaoui, 2011 [1]). The handling of realized ontologies becomes more and more difficult due to the significant increase number of new concepts to be integrated or updated.

Tourism is a complex science encompassing many professions:

(Accommodation, leisure, transport, sport, health...). Our aim is the re-conception of ontology OTM with the following ends:

- 1- Allowing definition of domain different point of view;

- 2- Providing consensual part shared by group of domain expert;
- 3- Allowing exchange and cooperation of experts in different domains;
- 4- Ensuring updating and evolution of global ontology.

The created ontology will be then hybrid (Wache et al., 2001[2]). It defines the semantics of each source of information in local ontology. A global “*consensual*” ontology encompasses common concepts among different sub-ontologies (domains). The main advantage of this representation is the simplicity to handle these local ontologies by experts. It contains only used concepts of their domains, whereas the consensual ontology allows experts to cooperate by using the definition links of matches setting.

## 2. METHODOLOGICAL APPROACH FOR SUB-ONTOLOGIES ALIGNMENT

OTM ontology realized in the framework of GECO-WES project is a consensual ontology composed of a set of sub-ontologies; each one of them describes a profession of tourism domain.

This ontology contains and shares all common concepts between different ontologies that interact with specific knowledge of this domain (tourism), as well as relations relating them. The treatment of the concepts included in the OTM ontology allows us to detect 20 sub-ontologies modeled profession in a legible tree form, easy to implement and simple to be evaluated by tourism experts (figure 1). The size of any sub-ontology varies according to the domain nature it describes.

<sup>1</sup> IRF-SIC : Forms Image Recognition – Intelligent Communication Systems.

<sup>2</sup> SIC: Intelligent Communication Systems.

<sup>3</sup> GECO-WES : Knowledge Management and WEB Semantic.

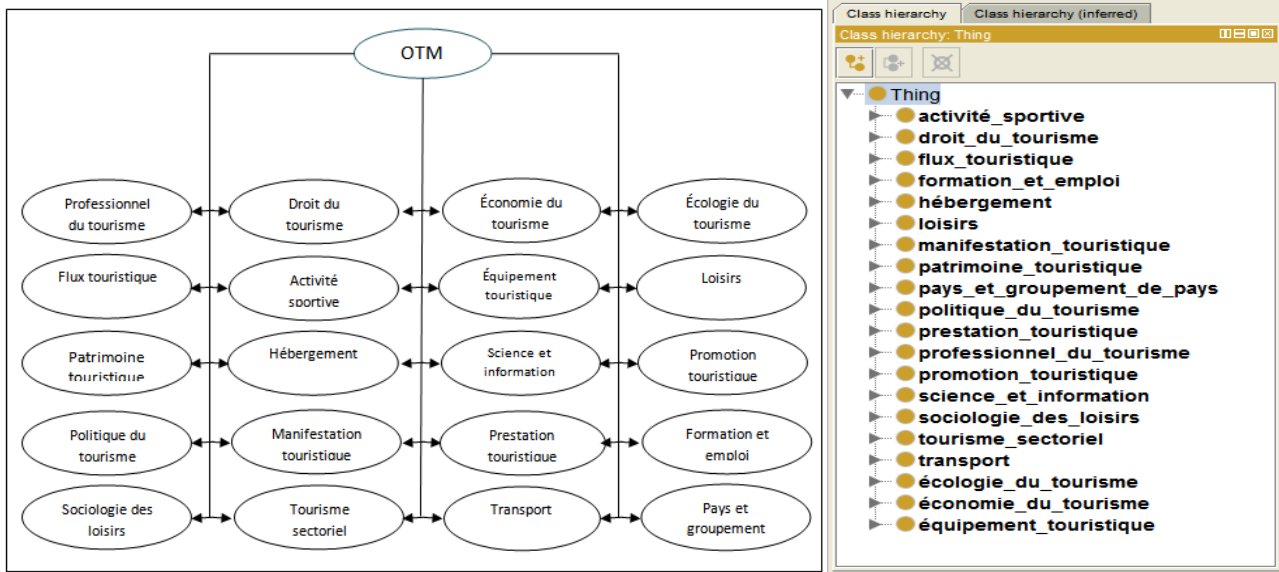


Fig.1: OTM Ontology Presentation

We thought of not making the example cumbersome by presenting a small part of ontologies (Figure 2). The sub-ontologies presented are: tourism accommodation and tourism benefits. The first one describes concepts and relations used by domain professionals in the process of

tourism accommodation such as (types of accommodation, mode of accommodation and accommodation capacity...).

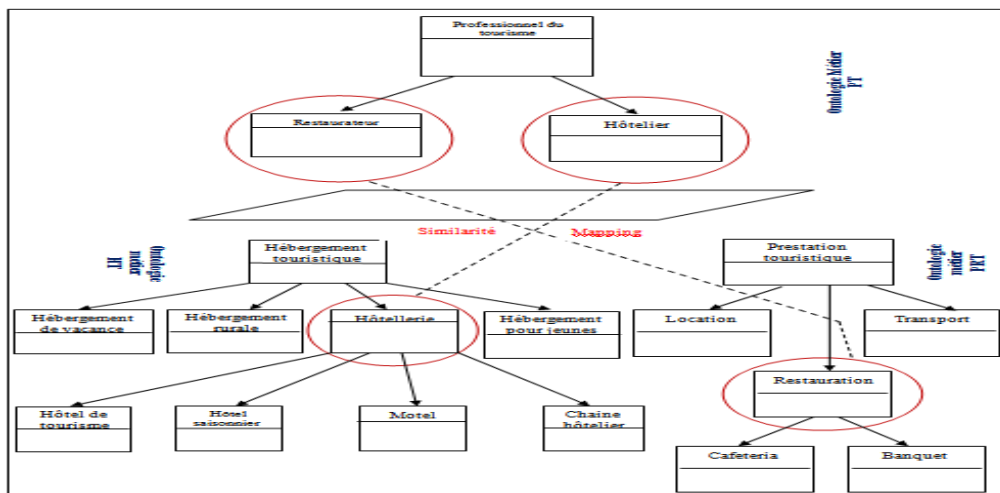


Fig.2: Example of similarities detected between the concepts belonging to local ontologies

The second one describes different types of benefits suggested to the tourist like (entertainment, location, commerce and travel...).

We present in this example only significant attributes as these concepts contain more than what is proposed.

To make these ontologies cooperate, we propose an approach with a methodological guide to ontologies designers in the process of creating new terminological and structural liaisons between concepts of domain ontology.

### 3. GENERAL ARCHITECTURE OF OUR APPROACH

There are now a set of tools and systems allowing the alignment of domain sub-ontologies. Among the best of them, we can mention those combining many measures to destruct similarity between two entities. However, there is no standard methodology that constitutes a unifying sustained framework to create mapping between sub-ontologies. We propose a methodological approach allowing the alignment of domain instable and evolving sub-ontologies by using matchers (terminological and structural). The whole performed treatment might be presented in seven stages as shown in the following figure.

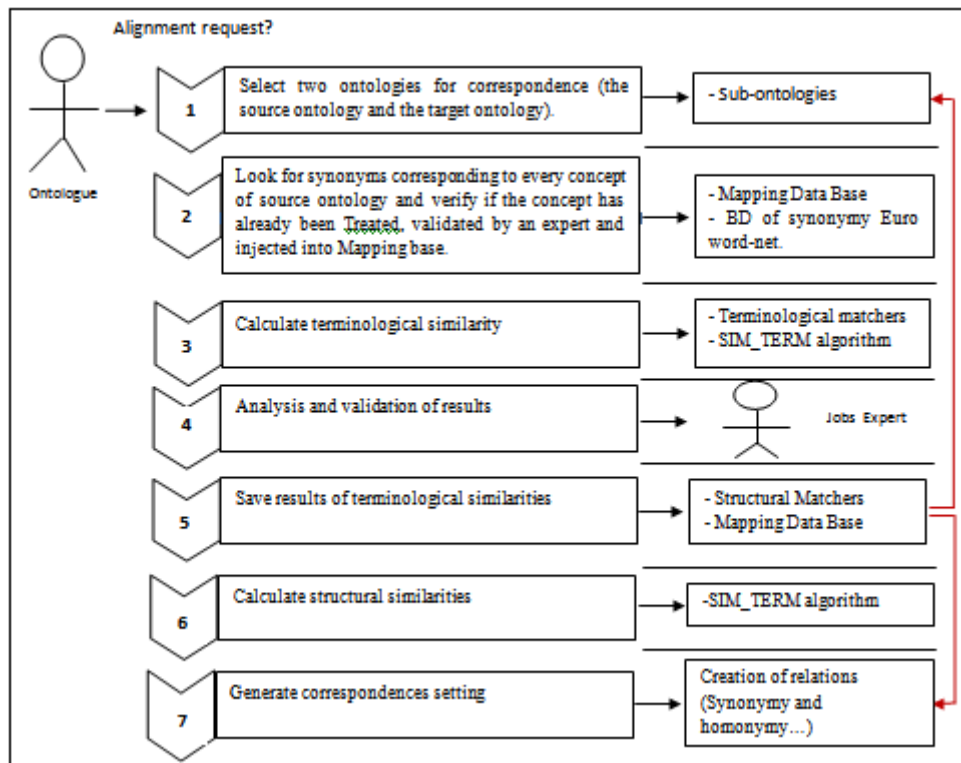


Fig 3: Stages of alignment of domain sub-ontologies.

To respect the norms of alignment, all ontologies correspondences to be set must be represented in OWL format.

The approach we suggest (Fig 3) will help designers of ontologies in the research process for similarities between distant ontological entities to set correspondences.

When an ontology researcher looks for similarities: **1/** he makes a request composed of the following elements (source ontology and target ontology). **2/** he verifies the existence of similarities (already calculated and validated in the BD mapping) corresponding to selected concept in the (stage 1), in order to reuse matches already treated and speed up time of execution. In case this concept has no entry in the BD mapping, it looks for synonyms equivalent to the entry concept in the BD of synonymy or in the descriptors lists (thesaurus, EurowordNet, Wolf...). **3/** calculates terminological and structural similarities measuring by using algorithms described in details in the following section. **4/** the results are a set of similarities between initial and found concepts in the research ontology. These results might be different types (equivalence, synonymy, subsumption...) which will be stored in the data base of similarities, then proposed to the domain expert to be validated and minimize link of incoherence. **5/** validated results of correspondence by expert are saved in the mapping base. **6/** calculation of structural similarities using specific algorithms with conceptual structure of graph and mathematical function declared in two structural matchers. **7/** the last stage enables

generating treated validated and approved correspondences by the domain expert.

#### 4. APPROACH RUNNING

The alignment process allows turning of correspondences setting between elements of source ontology towards target ontology that share common characteristics. The aim is to detect, and then create two types of the following links: (equivalence and specialization).

Many approaches of matching have been already developed (euzenat and shvaiko (2007) [3]), (kalfoglou (2003) [4]) and have even been implemented in the following domains (data base, semantic web, thesaurus...). To apply our approach, we propose treating the problematic of similarities detection (stage 3 figure 3) by developing matching techniques to detect structural and terminological similarities.

##### 4.1 Terminological techniques

Terminological methods allow the comparison of characters strings. Many studies have already been conducted in literature dealing with this problematic. As shown in the (figure 4), this process of comparison might be divided into three categories.

The first category is based on characters study. The second measure category is based on lexical analysis using Tokenization technique. The last category is based on the hybrid methods combining the two previous techniques.

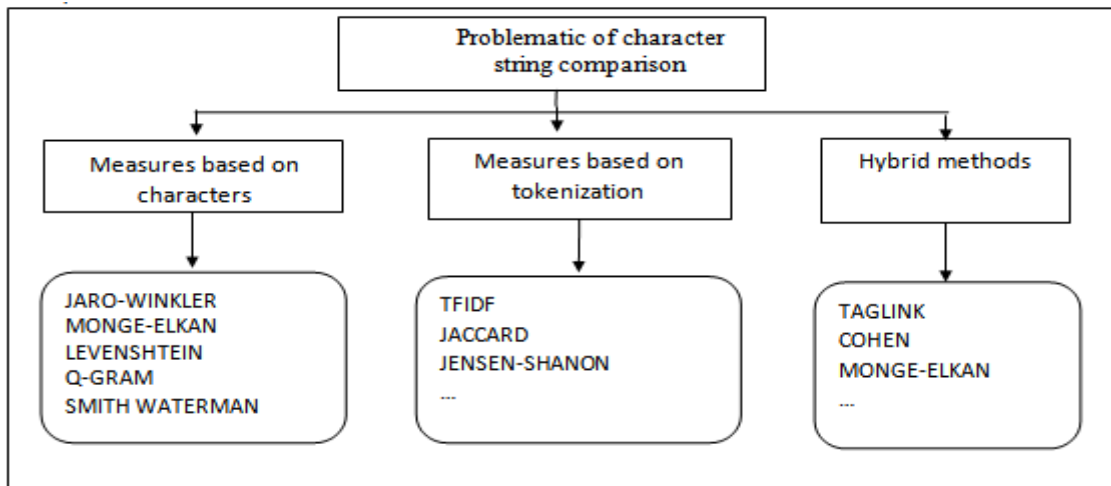


Fig. 4: Classification of methods of character string comparison.

All these measures might be applied to (name of entities, properties, labels, comments and instances) to detect similar ones. These solutions are drawn from the following hypothesis: two terms are similar, if they share enough elements in common (madche et al., 2002 [5]). Many ideas have been developed in literature using linguistic comparison of terms. To apply our approach, we suggest the use of terminological matchers:

- 1) Dictionary base matcher (Silva & Rocha 2003 [6]) which handles the lexical data base EuroWordNet by proposing for each word a list of SYNSETS corresponding to all listed meanings. Exploiting this base will allow us to identify types of relations between two concepts C1 and C2:

- $C1 \equiv C2$  if there is a meaning for C1 synonym of C2.
- $C1 \supseteq C2$  if there is a meaning for C1 hyperonymy of C2.
- $C1 \subseteq C2$  if there is a meaning for C1 hyponymy of C2.
- $C1 \perp C2$  if there is no relation between meanings of C1 and C2.

- 2) A matcher that allows terminological similarity calculation by mean of **Jaro-Winkler** function detecting existence of resemblances between characters strings (concepts).

The distance of Jaro  $D_j$  is defined by using the following formula:

**c1** and **c2** are two characters strings:

de Jaro distance between **c1** and **c2** is defined by :

$$D_j = 1/3 \left[ \left( \frac{m}{|c1|} \right) + \left( \frac{m}{|c2|} \right) + \frac{(m - t)}{m} \right]$$

Measure  $D_j \in [0,1]$ . Where zero represents lack of similarity

With:

$$\left\{ \begin{array}{l} |c_i| : \text{length of } c_i \text{ string.} \\ m : \text{number of characters corresponding to (i).} \\ t : \text{number of transposition (ii).} \end{array} \right.$$

(i) Two identical characters of  $c_{c1}$  et de  $c_{c2}$  are considered corresponding: If their distance (difference between their positions in the respective strings) does not exceed:

$$\left[ \frac{\max(|c1|, |c2|)}{2} \right] - 1$$

(ii) Number of transpositions is obtained by comparing i-ème character correspondent of  $c_1$  with i-ème character correspondent of  $c_2$ . Number of times these characters are different divided by two gives number of transpositions.

Fig. 5: function calculation of terminological similarities.

To develop our approach, we focus on the use of two matchers in a sequential way (Figure 5) within the algorithm (algorithm1). The latter allows the calculation of terminological similarities of concepts pairs of two ontologies. It takes two domain sub-ontologies entry from tourism

domain  $s_{o1}$  et  $s_{o2}$  to be aligned, a function  $F_{RECHERCHER\_SYN}$  (matcher 1) that looks in the dictionary EuroWordnet for different terminologies (synonymy, hyperonymy, hyponymy...) of a given concept. Also, the function  $\_SIM\_TERM$  (matcher 2) that allows

calculation of terminological similarity between two concepts (one concept belonging to target ontology and other one to source ontology) by the measure of JARO-WINKLER

adapted to characters strings like the one representing the concepts of our ontology OTM.

```

Algorithm: SIM TERM

Data:
S_O1: the source ontology.
S_O2: the target ontology.
F_RESEARCH_SYN: A function returning a list of (synonyms, Hyperonyms and
hyponyms) for a given concept.
F_SIM_TERM: A function which allows calculates terminological similarity by using
the distance of (Jaro-Winkler)
V_SC: Vector of the terminological links for a given concept
S_st: Structure of datum containing the information of similarity between two concepts.
Tab_st: list of terminological similarity.

Result:
Tab_v_sl: list of terminological similarity validated by the jobs experts

Begin:
/* Browse all concepts belonging to sub-ontology Source S_O1*/
For_each (C1 ∈ S_O1) {
    /* Search (synonyms, hyperonyms ...) of the selected concept */
    V_SC=F_RESEARCH_SYN(C1)

    /* Browse all concepts belonging to sub-ontology Target S_O2*/
    For_each (C2 ∈ S_O2) {
        For_each (S1 ∈ V_SC) {
            /* verify the existence of a similarity already calculate and validate with
            expert in database mappings */
            S_st = VERIFIER_BASE_MAP(c1, S1)

            IF S_sl Is_Empty THEN
                IF (S1.type == C2.type) THEN
                    /* Calculate the terminological similarity */
                    Sim_Term=CALCULSIMTERM(S1, C2, FC_SIM_TERM)
                    /* Add S1, C2 and Sim_Term to Tab_st */
                    Add((S1, C2, Sim_Term), Tab_st))
                ELSE
                    /* Add the similarity structure to Tab_st */
                    AJOUTER ((S_st, Tab_st)
                }
            }
        }

    /* The Results including into Tab_st must be validated by Job expert */
    Tab_v_sl = Validation_Results(Tab_st)

    /* inject the validated results into (DB de Mappings)*/
    Inject_Mapping (Tab_v_sl)
}
End

```

Algo1: Measure of terminological similarities.

## 4.2 Structural techniques

Structural techniques allow the detection of similarity between two entities according to structural information, where entities are semantically and syntactically forming a hierarchy or a graph of concepts. These structural techniques are divided into two families: (internal structural techniques and external structural techniques). The first category exploits only information describing the concepts attributes, whereas the second category treats the relations between these concepts.

### 4.1.1 Internal structural techniques

Structural techniques are based on the treatment of constraints (Rahm et Bernstein (2001)[7]). They use information of the attributes (field, cardinality of attributes, characteristics, transitivity of properties, restrictions) included in the internal structures of the concepts to calculate similarity.

### 4.1.2 External structural techniques

External structural techniques exploit existing relations between entities in a hierarchal structure. These relations contain subsumption relations (specialization or is\_a) or mereology (part-whole). The similarity between entities is determined according to their positions in their hierarchies.

Many ideas have been developed in literature to detect structural similarity between two distant entities. We propose the implement of two structural matchers.

- 1) The first matcher is sourced from the study realized in the framework of researches conducted by (Euzenat et Shvaiko (2006)[8]): According to this study: **“two concepts might be considered similar if and only”** if:

- Their **super-concepts** "father" are similar;
- Their **sub-concepts** "son" are similar;
- Their **"neighboring"** are similar.

- 2) The second structural matcher, allows calculation of structural similarity on the base of extracted and validated results in the stage of terminological similarity

measuring. The calculation of this measure is realized by similarity measure function ‘Match-Based Similarity’ developed by (Touzani et al. 2005

**C:** A concept belonging to source sub-ontology.  
**C':** A concept belonging to target sub-ontology.

**Tab\_vc:** Table of concepts (Neighboring) for a concept C selected in the sub-ontologie S\_O1.  
**Tab\_vc':** Table of concepts (Neighboring) for a concept C' selected in the sub-ontologie S\_O2.

**Gc:** set of concepts (Neighboring) belonging Tab\_vc.  
**Gc':** set of concepts (Neighboring) belonging Tab\_vc'.

**N.B:** Gc and Gc' contains concepts (Neighboring) of the same category.  
**Pc:** similarity weight of each category of concept.

$$\text{SIM\_STRUCT} = \sum_{(Gc, Gc') \in (\text{Tab\_vc}, \text{Tab\_vc}')} P_{c(Gc, Gc')} \times \text{FC\_Sim\_Str}(Gc, Gc')$$

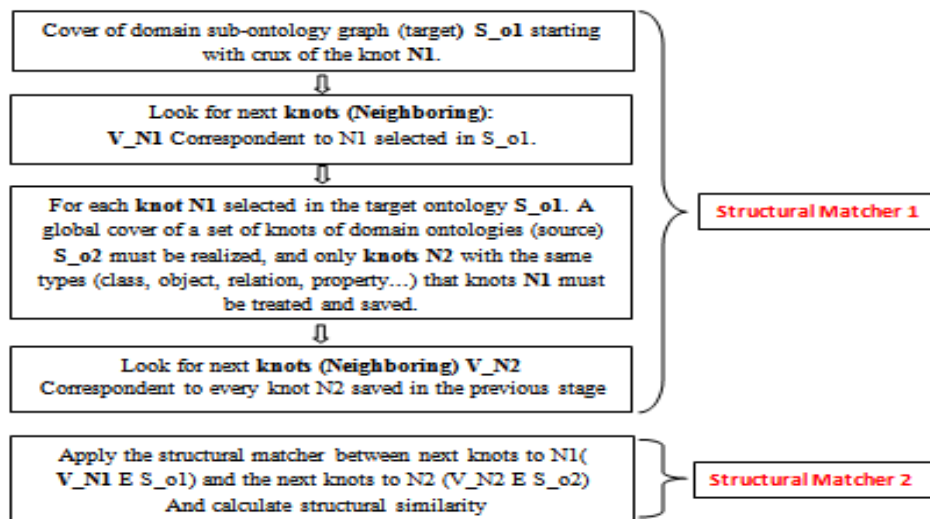
With:

$$\text{FC\_Sim\_Str}(Gc, Gc') = \sum_{(c, c') \in (Gc, Gc')} \text{Sim\_Term} / \text{Max}(|Gc|, |Gc'|)$$

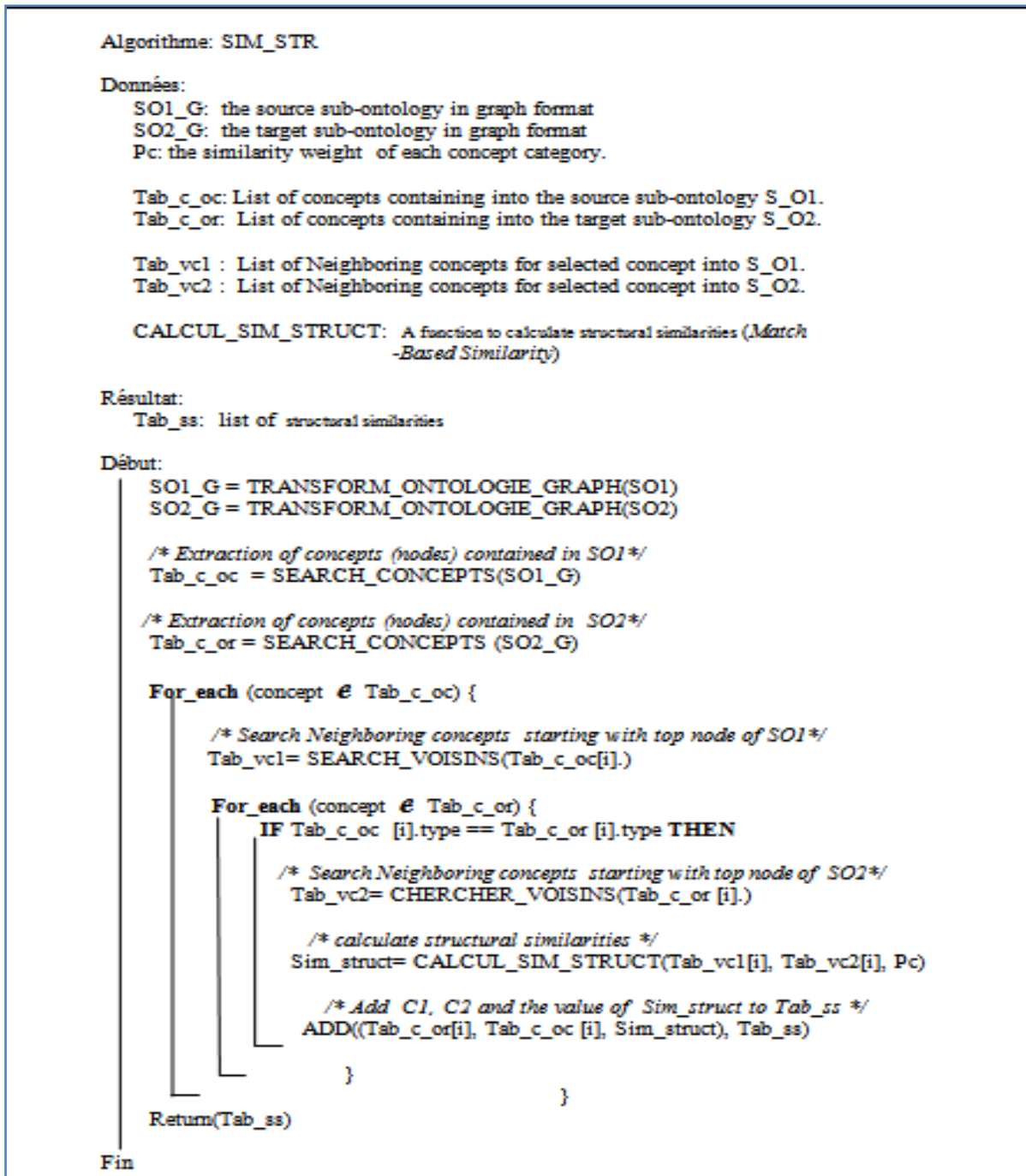
And  $\sum_{(Gc, Gc')} P_{c(Gc, Gc')} = 1$

Fig. 6: function calculation of structural similarities 'Match-Based Similarity'.

These two structural matchers are executed in a sequential way with the following order:



This set of chosen options to realize our approach to detect structural similarities are shown in details in the following algorithm:



Algo. 2: Measure of structural similarities.

## 5. CONCLUSION AND PERSPECTIVES

In this article, we have presented a methodological approach allowing the alignment of domain ontologies of tourism in Morocco. The alignment process proposed begins when ontology tries to detect possible liaisons between concepts of two local sub-ontologies. This setting of correspondence allows exchange and cooperation on different domain points of view of our domain study and ensures the evolution and maintenance of our global ontology.

As perspectives, we think of studying and developing semantic matchers based on the results of this research and analyzing impact of evolution of domain sub-ontologies on generated mapping

## 5. REFERENCES

- [1] S. Mouhim, A. Laoufi, C. Cherkaoui, E. Megder, D. Mammass, Marrakech 17-18 février 2011, "An Ontology based Knowledge Management System For Tourism", Information Systems And Economic Intelligence.
- [2] Wache H., Vogele T., Visser U. Stuckenschmidt H., Schuster G., Neumann H. Hubner S. Ontology-Based, Integration- A survey of Existing Approaches, IJCAI'01 Workshop on Ontologies and Information Sharing, 2001.

- [3] Jérôme Euzenat Pavel Shvaiko 2007, *Ontology Matching* Springer-Verlag New York, Inc. Secaucus, NJ, USA ©2007. ISBN:3540496114
- [4] Yannis Kalfoglou, Marco Schorlemmer 2003 *IF-Map: An Ontology-Mapping Method Based on Information-Flow Theory*. *Journal on Data Semantics I, Lecture Notes in Computer Science Volume 2800*, 2003, pp 98-127
- [5] Alexander Madche 2001. *F.C.A merge button-up merging of ontologies* In Proc 147 th seale( WA US) Pages 255-230.
- [6] Nuno Silva, Joao Rocha 2003. *Ontology Mapping for Interoperability in Semantic Web*. *Proceedings of the international Conference www/internet2003, Algarve, Portugal*.
- [7] Erhard Rahm, Philip A. Bernstein (2001), *A survey of approaches to automatic schema matching*. *The VLDB Journal* 10: 334–350.
- [8] Euzenat, M., Shvaiko, P., Stuckenschmidt, H., Svab, O., Svatek, V., van Hage, W.R., Yatskevich, M.: *Results of the ontology alignment evaluation initiative 2006*. In: *Proceedings of the First ESWC 2006 international workshop on ontology matching*. (2006).
- [9] Touzani M., « *Alignement des ontologies OWL-Lite* », Master's thesis, University of Montreal, 2005.