# Ontology Mapping
# Using Fuzzy Conceptual Graphs and Rules

Patrice Buche[1] and Juliette Dibie-Barthélemy[1,2] and Liliana Ibanescu[1]

[1]INRA Department of Applied Mathematics and Computer Science,
Mét@risk, 16 rue Claude Bernard, F-75231 Paris Cedex 05
[2]UFR Informatique, AgroParisTech, 16 rue Claude Bernard, F-75231 Paris Cedex 05
{Patrice.Buche,Juliette.Dibie,Liliana.Ibanescu}@agroparistech.fr

**Abstract.** This paper presents a new ontology mapping method between a source ontology and a target one considered as a reference. Both ontologies are composed of triplets of the form (object, characteristic, value). Values describing the objects of the reference ontology are hierarchically organized using the *a kind of* relation. The proposed method considers the ontology mapping problem as a rule application problem in the Conceptual Graph model. First, a vocabulary common to both ontologies is defined using mapping between values and characteristics. Each value of the source ontology is associated with a fuzzy set of values of the reference ontology. Then, the source ontology is translated into a fuzzy conceptual graph base and the reference ontology into a conceptual graph rule base. Finally, rules are applied into the fact base in order to find correspondences between objects of both ontologies. This method is implemented and applied to the mapping of ontologies in risk assessment in food products, and experimental results are presented.

## 1 Introduction

Information systems which are characterized by the presence of multiple and independent knowledge representation are concerned by the problem of the interoperability among them. Mappings play a key role to treat that problem and may be used for different purposes (schema or ontology integration, ontology engineering, ...). Ontology matching is defined as a process that takes two ontologies as input and returns a mapping which identifies corresponding concepts in the two ontologies by taking into account their descriptions and constraints in terms of names, properties and semantic relations. The problem on the ontology matching problem has been widely investigated in the literature (see [5, 8, 7, 2]).

In the framework of Conceptual Graphs (CG), previous works [6] have shown that this model can be extended to ontology matching based on conceptual properties. In this paper, we want to use the CG model when ontology matching is based simultaneously on lexical and conceptual properties. More precisely, we want to address the mapping process of a source ontology with a target ontology considered as a reference. Both ontologies are composed of triplets of the form (object, characteristic, value). There is no class categorization for objects and characteristics, and the values contained in the reference ontology are organized according to the *a kind of* partial value function. We propose to use fuzzy CGs [10] to represent and to match ontologies for three main reasons: (i) the support of the CG model is well adapted to the representation of the taxonomies of the reference ontology; (ii) the projection operation takes into account the specialization relation between values of the ontologies; (iii) the fuzzy extension encodes similarities between values and objects of the ontologies.

The aim of the proposed mapping method is to establish correspondences between objects of two ontologies. The mapping problem adressed in this paper is not a symmetric problem since one of the two ontologies is considered as a reference. So we propose a new ontology mapping method in which the reference ontology is considered as a rule base and the source ontology as a fact base. The ontology mapping problem then becomes a rule application problem. Nevertheless, in order to apply rules into a fact base, both rules and facts must be defined with the same vocabulary. So, our mapping method can be divided into three main steps. The first step (section 2) consists in defining a vocabulary common to the source and the reference ontologies. The second step (section 3) concerns the translation of the source ontology into a fact base and of the reference ontology into a rule base. The third step (section 4) deals with the application of the rules into the fact base in order to find correspondences between objects of both ontologies. Finally, experimental results are presented in section 5.

## 2 Definition of a common vocabulary

We have chosen the Conceptual Graph (CG) model as formalized in [1] in order to represent and to compare objects of a source ontology denoted $\mathcal{S}$ with objects of a reference ontology denoted $\mathcal{R}$. The CG model contains (i) the terminological knowledge made of a concept type lattice which contains a smallest type denoted $\bot$ and a biggest one denoted $\top$, a relation type set possibly organized in hierarchy, a set of individual markers enabling the designation of instances and a conformity relation between markers and types, (ii) a CG fact base built on the terminological knowledge and (iii) rules of the form $G_H \Rightarrow G_C$ where $G_H$ represents the hypothesis of the rule and $G_C$ its conclusion.

In order to compare objects of $\mathcal{S}$ with objects of $\mathcal{R}$, we would like to use the projection operation on CGs. But the objects of $\mathcal{S}$ are not defined with the same vocabulary as the objects of $\mathcal{R}$. Since the ontology $\mathcal{R}$ is a reference one, we propose to express each object of $\mathcal{S}$ in terms of characteristics and values of $\mathcal{R}$. For that, we define a mapping between values and characteristics of $\mathcal{S}$ and $\mathcal{R}$. We only briefly recall this mapping which has already been presented in [3].

First, each value $v$ of $\mathcal{S}$ is associated with a set of values $\{w_1, \ldots, w_n\}$ of $\mathcal{R}$, weighted by their lexical closeness to the value $v$ using the Dice coefficient. Such a set of values is represented by a fuzzy set [11, 12].

**Example 1** *Let* pollock raw *be a value of* $\mathcal{S}$*. Let* pollock, Alaska pollock *be values of* $\mathcal{R}$. $\mu_{pollockraw}$=\{ *0.66/pollock + 0.5/Alaska pollock* \}.

The lexical mapping between values is used to identify correspondences between characteristics of $\mathcal{S}$ and $\mathcal{R}$. The result of the mapping between values and characteristics of $\mathcal{S}$ with values and characteristics of $\mathcal{R}$ is defined below.

**Definition 1** *We call* linked values *of the source ontology* $\mathcal{S}$*, denoted* $LV_{\mathcal{S}}$*, the set of values of* $\mathcal{S}$ *such that each of them is associated with a set of values of the reference ontology* $\mathcal{R}$ *with a given relevance score, represented by a discrete fuzzy set. We call* linked characteristics *of* $\mathcal{S}$*, denoted* $LC_{\mathcal{S}}$*, the set of characteristics of* $\mathcal{S}$ *such that each of them is associated with one characteristic of* $\mathcal{R}$*.*

Thanks to this mapping, we can now present the terminological knowledge common to $\mathcal{S}$ and $\mathcal{R}$. The concept type set is composed of the object names of $\mathcal{S}$ and $\mathcal{R}$, the set of characteristics of $\mathcal{R}$, the hierarchized set of values of $\mathcal{R}$ and the concept type *NumVal*. The relation type set is composed of the relation types *HasForCharac*, *HasForValue*, *IsAnnotatedBy* and *HasForScore*. The set of individual markers contains values of the reference domain of the real numbers $\mathbb{R}$.

## 3 Translation of the ontologies into fact and rule bases

Since the vocabulary common to the source ontology $\mathcal{S}$ and the reference ontology $\mathcal{R}$ has been defined, we can now deal with the second step of our mapping method i.e. to translate $\mathcal{S}$ into a CG fact base and $\mathcal{R}$ into a CG rule base.

### 3.1 Translation of the source ontology into a fuzzy CG base

Each object of $\mathcal{S}$ is represented by a CG using the terminological knowledge described above: each of its characteristics and each of its associated values are represented by means of their corresponding characteristic and values in $\mathcal{R}$. Since each value of the object in $\mathcal{S}$ is associated with a fuzzy set of values in $\mathcal{R}$[1], the CG contains fuzzy values. We have proposed in [10] an extension of the CG model to represent fuzzy values: a fuzzy set with a hierarchized reference domain can be represented in a concept vertex as a *fuzzy type*.

**Definition 2** *Let* $f$ *be the fuzzy value function which associates each value of* $LV_{\mathcal{S}}$ *with its corresponding values in the reference ontology* $\mathcal{R}$ *and their relevance score. Let* $g$ *be the value function which associates each characteristic of* $LC_{\mathcal{S}}$ *with its corresponding characteristic in* $\mathcal{R}$*. Let* $C_T = \{ Charac_1^T, \ldots, Charac_p^T \}$ *be the set of characteristics of* $\mathcal{R}$*. Let* $Name_Y$ *be the name of an object* $Y$ *of the source ontology* $\mathcal{S}$*. Let* $C_Y^T = \{ g(Charac_1^Y), \ldots, g(Charac_m^Y) \} \in C_T$*,* $m \leq p$*, be the set of characteristics associated with* $Y$ *in* $\mathcal{R}$*, where* $Charac_i^Y \in LC_{\mathcal{S}}$*,* $i \in [1, m]$*. Let* $C'_T = \{ Charac_l^T, \ldots, Charac_k^T \}$*,* $p - m \leq l \leq k \leq p$*, be the set of characteristics of* $\mathcal{R}$ *such that* $C'_T = C_T \setminus C_Y^T$*. Let* $Value_1, \ldots, Value_m$ *be the values associated with the characteristics of* $Y$ *and belonging to* $LV_{\mathcal{S}}$*. Then each object* $Y$ *of* $\mathcal{S}$ *can be represented by the CG* $G_Y^T$ *of Figure 1.*

[1] This fuzzy set of values has a semantic of similarity and represents the ordered list of the most similar values of $\mathcal{R}$ associated with a value of $\mathcal{S}$.

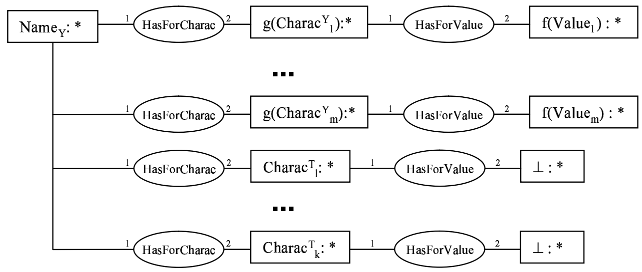


**Fig. 1.** The CG $G_Y^T$ associated with an object $Y$ of $\mathcal{S}$.

**Example 2** *Let* fresh fish *be an object of* $\mathcal{S}$. *Its associated list of couples (characteristic : value) is: (presentation: whole) and (which fish ?: pollock raw). Figure 2 presents the CG* $G_{ff}^T$ *associated with* fresh fish, *where g(presentation)= 'origin of main ingredient' and g(which fish?)= 'physical state, shape'.*

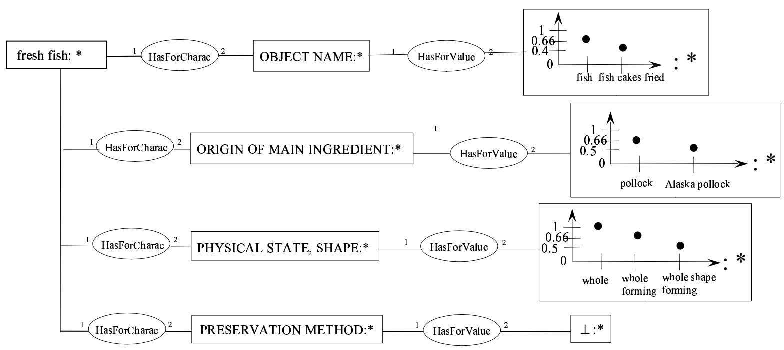


**Fig. 2.** The CG $G_{ff}^T$ associated with the object **fresh fish** of $\mathcal{S}$.

### 3.2 Translation of the reference ontology into CG rules

Each object of $\mathcal{R}$ is represented by means of a CG rule which allows objects of $\mathcal{S}$ to be annotated with objects of $\mathcal{R}$ according to the correspondences between their characteristics and associated values.

**Definition 3** *Let* $C_T = \{ Charac_1^T, \ldots, Charac_p^T \}$ *be the set of characteristics of the reference ontology* $\mathcal{R}$. *Let* $Name_X$ *be the name of an object* $X$ *of* $\mathcal{R}$. *Let* $C_T^X = \{ Charac_1^X, \ldots, Charac_n^X \}$, $n \leq p$, *be the set of characteristics associated with* $X$. *Let* $C'_T = \{ Charac_l^T, \ldots, Charac_k^T \}$, $p - n \leq l \leq k \leq p$, *be the set of characteristics of* $\mathcal{R}$ *such that* $C'_T = C_T \setminus C_X^T$. *Let* $Value_1, \ldots, Value_n$ *be the values associated with the characteristics of* $X$. *Then, each object* $X$ *of* $\mathcal{R}$ *can* *be represented by the CG rule $R_X$ of Figure 3 where the marker *scoreproj is the adequation degree between the hypothesis of $R_X$ and a CG into which there exists a $\delta$-projection (see Definition 4).*

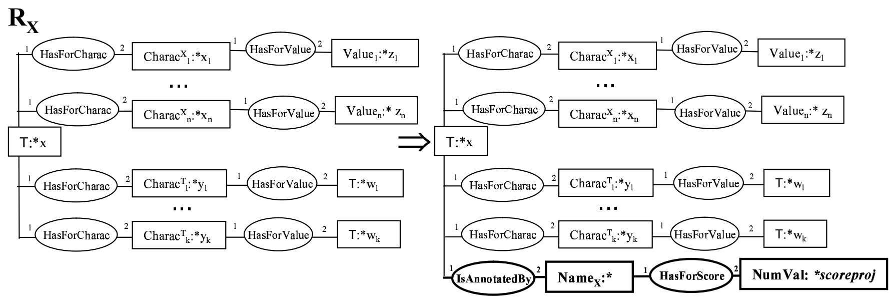


**Fig. 3.** The CG rule $R_X$ associated with an object $X$ of $\mathcal{R}$. Vertices framed in bold correspond to the conclusion of the rule.

**Example 3** *Let* cod, raw *be an object of $\mathcal{R}$. Its associated list of couples (characteristic : value) is: (origin of main ingredient: cod or codfish), (physical state, shape: whole, shape solid) and (preservation method: preserved by refrigeration). Figure 4 presents the CG rule associated with the object* cod, raw *of $\mathcal{R}$.*

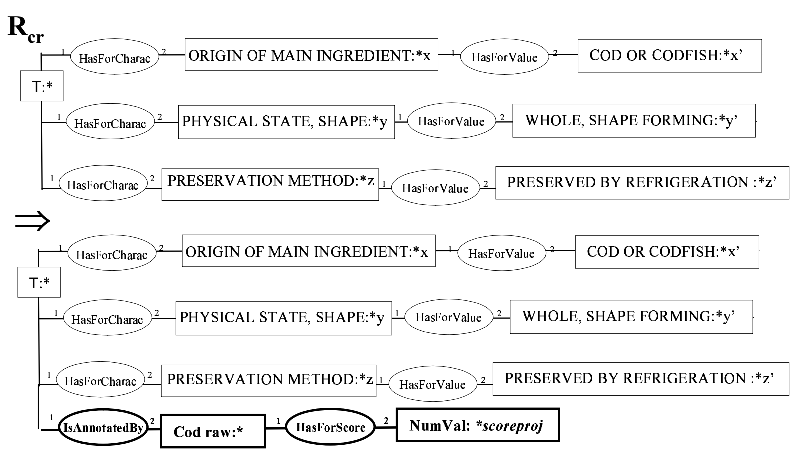


**Fig. 4.** The CG rule $R_{cr}$ associated with the object *cod, raw* of $\mathcal{R}$.

## 4 Using CG rules for fuzzy matching of objects

Objects of the ontologies $\mathcal{S}$ and $\mathcal{R}$ are now represented by comparable CGs using the same vocabulary. The objects of $\mathcal{S}$ are represented by fuzzy CGs and the objects of $\mathcal{R}$ by CG rules. The next and last step of our mapping method consists in applying the CG rules into the fuzzy CGs in order to find correspondences between objects of $\mathcal{S}$ and objects of $\mathcal{R}$. These rules application allows the objects of $\mathcal{S}$ to be enriched with annotations that are sets of similar objects of $\mathcal{R}$.

The rule application requires to compare a crisp CG which represents the hypothesis of a rule with a fuzzy CG which represents an object of $\mathcal{S}$ and may contain fuzzy values. This comparison is made using the $\delta$-projection which is an extension of the projection operation defined in [10, 3].

**Definition 4** *A* $\delta$-projection *from a crisp CG $G$ into a fuzzy CG $G'$ is a triple ($g$, $h$, $\delta$), $g$ (resp. $h$) being a mapping from the set of concept (resp. relation) vertices of $G$ into the set of concept (resp. relation) vertices of $G'$ such that: (i) the edges and their numbering are preserved; (ii) the labels of the relation vertices may be restricted; (iii) $\forall$ crisp concept vertex $c_i \in G$, $i \in [1, \ldots, n]$, labelled $t_i : m_i$, $c_i$ is mapped with its image $g(c_i) \in G'$ labelled $t'_i : m'_i$, with an adequation degree $\delta_i = \mu_{clos(t'_i)}(t_i)$, $\mu_{clos(t'_i)}$ being the membership function of the fuzzy type closure of $t'_i$. The adequation degree of $G$ by $G'$ is $\delta = min_{i=1,\ldots,n}\delta_i$.*

We can now identify the correspondences between objects of $\mathcal{S}$ and $\mathcal{R}$. Each rule associated with each object of $\mathcal{R}$ is $\beta$-applied into the fuzzy CGs representing the objects of $\mathcal{S}$, $\beta$ being a threshold allowing the end-user to avoid too bad correspondences between objects. The $\beta$-application is an extension of the rule application defined in [9].

**Definition 5** *There exists a* $\beta$-application *from a rule $G_H \Rightarrow G_C$ into a CG $G$ if there exists a $\delta$-projection from $G_H$ into $G$ such that $\delta \geq \beta$.*

**Example 4** *Let us consider the object* fresh fish *of $\mathcal{S}$ described in Example 2 and represented by the CG $GT_{ff}$ of Figure 2. Let us consider the object* cod, raw *of $\mathcal{R}$ described in Example 3 and represented by the rule $R_{cr}$ of Figure 4. There exists a* 0.5*-projection from the hypothesis of $R_{cr}$ into $GT_{ff}$. So, $R_{cr}$ can be* 0.4*-applied into $GT_{ff}$. The resulting CG $R[GT_{ff}]$ is described in Figure 5.*

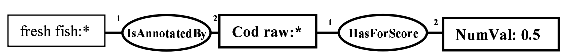


**Fig. 5.** The resulting CG $R[GT_{ff}]$ obtained from the application of the rule $R_{cr}$ from Figure 4 into the CG $GT_{ff}$ from Figure 2 is partially shown here. It includes the one of Figure 2 completed by the annotation framed in bold of this figure.

Thus, at the end of this mapping process, each object $Y$ of the source ontology $\mathcal{S}$ is associated with a set of candidate objects (see Definition 6) of the reference ontology $\mathcal{R}$, weighted by their adequation degrees to the object $Y$.

**Definition 6** *An object* $X$ *of the reference ontology* $\mathcal{R}$ *is a* candidate *for an object* $Y$ *of the source ontology* $\mathcal{S}$ *with the adequation degree* $\delta$ *if the generic concept vertex of type* $Name_Y$ *is linked by the relation vertex* $IsAnnotatedBy$ *to the generic concept vertex of type* $Name_X$ *which is linked by the relation vertex* $HasForScore$ *to the individual concept vertex (*$NumVal$*:* $\delta$*).*

**Example 5** *According to Example 4, the object* cod, raw *of* $\mathcal{R}$ *is a candidate for the object* fresh fish *of* $\mathcal{S}$ *with the adequation degree* 0.5*.*

## 5 Experimentation

We have developed methods to estimate the exposure of a given population of consumers to chemical contaminants using two databases: the first one, called *CONTA*, considered as the reference ontology $\mathcal{R}$, gives the degree of chemical contamination for 472 food products; the second one, called *CONSO*, considered as the source ontology $\mathcal{S}$, stores household purchases of 2595 food products.

We have realised an expert manual mapping: 398 food products from the *CONSO* ontology (i.e. 84.32% from 472) have been associated with 2041 food products from the *CONTA* ontology (i.e. 78.65% from 2595) by 3258 mappings. Only 118 mappings (i.e. 3.82% from 3258) associate one food product from the *CONSO* ontology with exactly one food product from the *CONTA* ontology.

Table 1 gives precision (the percent of the found correct mappings to the found mappings) and recall (the percent of the found correct mappings to the correct mappings found manually) for different correspondences. Mapping of food product names, without mapping of characteristics, permits to retrieve half of the manual matches but has a very bad precision (8.8%). Mapping of characteristics enhances the recall till around 77%. We have also evaluated the influence of the taxonomy defined on the values of $\mathcal{R}$: for the mapping of 6 (resp. 20) characteristics, 6.96% (resp. 8.38%) of 74.40% (77.04%) are obtained.

| *#nb charac* | $\#found$ | $\#correct$ | $p \times 100$ | $r \times 100$ |
|---|---|---|---|---|
| 0 | 18 283 | 1 608 | 8.80 | 49.36 |
| 6 | 72 365 | 2 424 | 3.34 | 74.40 |
| 20 | 120 468 | 2 510 | 2.08 | 77.04 |

**Table 1.** Results obtained with a number of mapped characteristics from 0 to 20

## 6 Conclusion

In this paper we present an ontology mapping method between a source ontology and a reference one. Both ontologies are composed of triplets of the form (object, characteristic, value). Values describing the objects of the reference ontology are hierarchically organized using the *a kind of* relation. First, all the objects of the source ontology $\mathcal{S}$ are represented into a fuzzy CG base, denoted ${\mathcal{KB}_\mathcal{S}}^T$. Then, all the objects of the reference ontology $\mathcal{R}$ are represented as a set of CG rules, denoted $Rules_{\mathcal{R}}$. Finally, the rules from $Rules_{\mathcal{R}}$ are applied into ${\mathcal{KB}_{\mathcal{S}}}^T$. This application produces an annotation for objects from $\mathcal{S}$ that encodes correspondences with objects from $\mathcal{R}$ and the associated adequation degrees. We have shown in this paper that, thanks to our fuzzy extension of the CG model, it is possible to represent and manipulate lexical mapping results combined with semantic properties. This method has been implemented and applied to the mapping of ontologies in risk assessment in food products. Our experimentation shows that the method has a rather good recall but a poor precision.

A first perspective to enhance our method is to study other comparaison techniques between characteristics and values such as semantic techniques or contextual matching techniques. An other perspective is to apply, in post-treatement, semantic constraints on the generated mappings between objects. Finally, we want to compare our results with the one obtained using other ontology alignement methods thanks to ontology alignment comparison systems ([4]).

## References


1. J. F. Baget and M. L. Mugnier, *Extensions of simple conceptual graphs: the complexity of rules and constraints*, Journal of Artificial Intelligence Research **16** (2002), 425–465.
2. S. Castano, A. Ferrara, S. Montanelli, G. N. Hess, and S. Bruno, *BOEMIE (Bootstrapping Ontology Evolution with Multimedia Information Extraction), State of the Art on Ontology Coordination and Matching*, FP-027538 Deliverable 4.4, 2007.
3. D. Doussot, P. Buche, J. Dibie-Barthélemy, and O. Haemmerlé, *Using Fuzzy Conceptual Graphs to Map Ontologies*, (2006), 891–900.
4. Jérôme Euzenat, *An api for ontology alignment*, International Semantic Web Conference, Lecture Notes in Computer Science, vol. 3298, 2004, pp. 698–712.
5. Jérôme Euzenat and Pavel Shvaiko, *Ontology Matching*, Springer, 2007.
6. Frédéric Fürst and Francky Trichet, *Heavyweight ontology matching - a method and a tool based on the conceptual graphs model*, ICEIS 2007, 2007, pp. 265–270.
7. Yannis Kalfoglou and Marco Schorlemmer, *Ontology Mapping: The State of the Art*, Semantic Interoperability and Integration (Y. Kalfoglou, M. Schorlemmer, A. Sheth, S. Staab, and M. Uschold, eds.), Dagstuhl Seminar Proceedings, no. 04391, Internationales Begegnungs- und Forschungszentrum fuer Informatik (IBFI), Schloss Dagstuhl, Germany, 2005.
8. Natalya F. Noy, *Semantic Integration: A Survey Of Ontology-Based Approaches*, SIGMOD Record Special Issue on Semantic Integration **33** (2004), no. 4.
9. E. Salvat and M.L. Mugnier, *Sound and complete forward and backward chainings of graph rules*, Proceedings of the 4th International Conference on Conceptual Structures, ICCS'96, Lecture Notes in Artificial Intelligence 1115, Springer-Verlag (Sydney, Australia), August 1996, pp. 248–262.
10. R. Thomopoulos, P. Buche, and O. Haemmerlé, *Different Kinds of Comparisons Between Fuzzy Conceptual Graphs*, Proceedings of the 11th International Conference on Conceptual Structures, ICCS'2003, Lecture Notes in Artificial Intelligence #2746 (Dresden, Germany), Springer, July 2003, pp. 54–68.
11. L. Zadeh, *Fuzzy sets*, Information and control **8** (1965), 338–353.
12. ______, *Fuzzy sets as a basis for a theory of possibility*, Fuzzy Sets and Systems **1** (1978), 3–28.