Clustering of conceptual graphs with sparse data

GANASCIA J.-G. and VELCIN J.

LIP6, Université Paris VI
8 rue du Capitaine Scott
75015 PARIS
FRANCE
{jean-gabriel.ganascia,julien.velcin}@lip6.fr

Abstract. This paper gives a theoretical framework for clustering a set of conceptual graphs characterized by sparse descriptions. The formed clusters are named in an intelligible manner through the concept of stereotype, based on the notion of default generalization. The cognitive model we propose relies on sets of stereotypes and makes it possible to save data in a structured memory.

Introduction

This paper proposes a formal approach for the clustering of conceptual graphs characterized by sparse descriptions. According to R. Michalski [1], conceptual clustering is a learning task that takes a set of object descriptions as input and creates a classification scheme. In addition, not only does conceptual clustering have to cluster facts, but it also has to name these clusters. It is crucial to build intelligible descriptions of all generated clusters that can summarize data in an understandable manner. The facts we have considered are rough descriptions of situations taken, for instance, from newspaper articles. These data are not given in an array where all the predefined attributes have been given values, but are characterized by heterogeneous descriptions. In order to deal with such types of data, it was necessary to extend clustering techniques to sparse descriptions.

The names of the clusters, also called labels, are usually calculated either using classical generalization (cf. UNIMEM [2]) or probabilistic generalization (cf. COBWEB [3]). This means that these names are derived from properties common to the facts belonging to the clusters, but in the case of sparse data it fails because there are very few common descriptors. This paper proposes a new approach that could deal with such cases, by building cluster names made up of all the non-contradictory features belonging to the clusters. These names are called stereotypes, by analogy with the concept of prototypes, which are based on a new notion of generalization, default generalization. This concept of stereotypes is well adapted to sets of facts described using sparse descriptions. We therefore consider a cognitive model that structures memory from facts into sets of stereotypes.

Section 1 presents this cognitive model. The first subsection introduces the notion of default generalization – and shows how it is linked to the compatibility relation in the conceptual graph formalism – which is the basis of our model. The second subsection describes the concept of stereotypes and compares it with the well-known concept of prototypes. Section 2 introduces tools and strategies to build the sets of stereotypes, and section 3 examines the question of validation and suggests an application.

1 The cognitive model

Memory has to be structured in order to retrieve old information without registering all the facts and to infer new knowledge from these recorded data. The notion of default generalization will be presented first, followed by the concept of stereotypes.

1.1 Notion of default generalization

The facts we have considered have a large number of missing values, as a result of which much of the information has to be guessed in order to perform the clustering. R. Reiter presents in [4] a logic for default reasoning where default rules enable new formulas to be inferred if the hypotheses are not inconsistent. The following rule translates a possible reasoning for the end of the 19th in France:

`politician(X) ∧ introducedAbroad(X) : ¬ diplomat(X) / traitor(X)`

This statement means that a politician who is introduced abroad is a traitor towards his own country if we cannot prove that he is a diplomat. Here, a stereotype stored in the memory will complete a fact if this fact has no contradictory features and is not more similar to another stereotype. The second condition can be checked using the dissimilarities measure and will be seen in the next section. The first and most important condition is verified if the stereotype *generalizes* the fact *by default*. This means that the fact can be completed in such a way that it can now be generalized by the stereotype.

Let us consider the graph g associated with a fact in which there is a very large number of missing values. The missing values can be guessed and completed to obtain a more specific graph g_S. Let us note that we follow here the notations given by Sowa in [5] (definition 3.5.1): $g_S \leq g$ means that g_S is a specialization of g and g is a generalization of g_S [1].

Now, let s be one stereotype belonging to the structured memory. If this stereotype is more general than g_S, ie $g_S \leq s$, then it generalizes g *by default*. More formally:

Definition 1. *Let f be a fact and s a stereotype, both of them represented using conceptual graphs.* s generalizes f by default *if and only if there exists a graph*

[1] Many other formalizations exist that encompass the notion of generalization, for instance least general generalization [6] or, more recently, Inductive Logic Programming [7]. For the sake of clarity, we will limit ourselves here to just one formal framework, that designed by Sowa.

g_S with $g_S \leq f$ and $g_S \leq s$. g_S is therefore a graph formed by the join operator performed on the graphs f and s.

Property 1. The notion of default generalization is more general than that of classic generalization. Let g and g' be two conceptual graphs. If g generalizes g' then g generalizes g' by default.

Property 2. The default generalization is a symmetrical relation. Let u_1 and u_2 be two conceptual graphs. If u_1 generalizes u_2 by default, then u_2 generalizes u_1 by default too.

Fig. 1 presents the fact *The meal of Jules is composed of steak, red wine, and ends with a cup of very hot coffee* which can be generalized by default by the stereotype *The meal is composed of steak with potatoes and French bread, and ends with a cup of coffee* because the fact can be completed to *The meal of Jules is composed of steak with potatoes and French bread, red wine, and ends with a cup of very hot coffee.* If the stereotype had presented a meal ending with a liqueur, it would not match the fact and so could not generalize it by default.

This notion allows us to propose stereotypes that label sets of facts and not generalize each fact in the ordinary way but using a default reasoning. Fig. 3 shows three facts that could only be generalized in the ordinary way by the graph `[MEAL]→(mainly composed of)→[DISH]`, which does not characterize the observations satisfactory. Fig. 4 presents a stereotype that can cover these facts with the default generalization notion.

Let us now compare in more detail the concepts of stereotype and prototype.

1.2 Concept of stereotype

Eleanor Rosch saw the categorization itself as one of the most important issues in cognitive science. She observed that children learned how to classify first in terms of concrete cases rather than through defining features. She therefore introduced the concept of prototype [8] as the ideal member of a category. Members of the same class might share only a few of those features but are closer to the same prototype, and so are grouped together. For example, a robin is closer to the bird prototype than an ostrich, but they are both closer to it than they are to the prototype of a fish, so we call them both birds. However, it takes longer to say an ostrich is a bird than it takes to say a robin is a bird, because the ostrich is further from the prototype.

Sowa defines a prototype in [5] as a typical instance formed by joining one or more schemata. Instead of describing a specific individual, it describes a typical or "average" individual. Fig. 2 shows the example of a prototype for ELEPHANT in linear form.

Many sufficiently complete observations have to be considered in order to construct such an individual. We therefore propose a new concept, stereotype, which is quite close to that of prototype but more adapted to missing values. A stereotype is an imaginary fact that combines features found in the facts covered

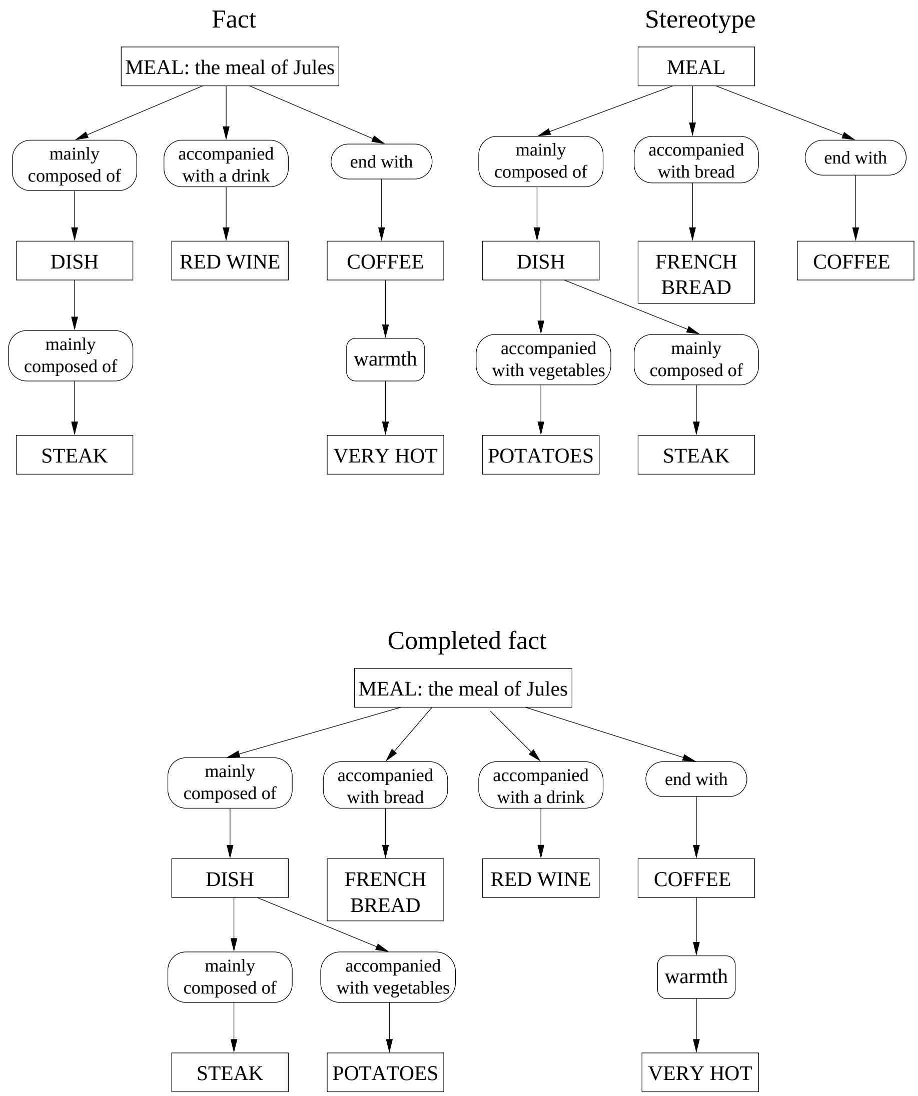

Fig. 1. The stereotype *generalizes by default* the presented fact. The fact below is the result of the join operator.

```
prototype for ELEPHANT(x) is
    [ELEPHANT: *x]-
        (CHRC)→[HEIGHT:@3.3m]
        (CHRC)→[WEIGHT:@5400kg]
        (COLR)→[DARK-GRAY]
        (PART)→[NOSE]-
                (ATTR)→[PREHENSILE]
                (IDNT)→[TRUNK],
        (PART)→[EAR:*]-
                (QTY)→[NUMBER:2]
                (ATTR)→[FLOPPY],
        (PART)→[TUSK:*]-
                (QTY)→[NUMBER:2]
                (MATR)→[IVORY],
        (PART)→[LEG:*]→(QTY)→[NUMBER:4]
        (STAT)→[LIVE]-
                (LOC)→[CONTINENT:Africa | Asia]
                (DUR)→[TIME:@50 years].
```

Fig. 2. A prototype for ELEPHANT.

rather than by performing averages. There is no contradiction between a fact and the related stereotype which covers this fact and may be used to guess missing data. It is because calculating an average is not adapted to a large number of missing values, especially when they are symbolic, that we have introduced the concept of stereotype which is built up from facts just like the pieces of a puzzle.

Fig. 4 represents a stereotype formed from three facts taken from fig. 3 corresponding to French meals. The missing values in these facts can be completed using default reasoning with the corresponding values in the stereotype because there is no contradiction between them. Thus, we can infer for instance that Tom drinks wine or that the Petit family eat French bread with their meal.

2 Construction of the stereotypes

This section is intended to show a way to automatically structure the memory from a set of facts. The main objective of our work is to find a set of stereotypes that summarize observed situations. The associated graphs have to contain enough information not only to classify new observations but also to make possible the reasoning on those situations.

2.1 The dissimilarities measure

We need to establish a measure to calculate the distance between two conceptual graphs, one being a fact to be covered and the other the candidate-stereotype. Previous work deals with graph matching and an interesting method to calculate

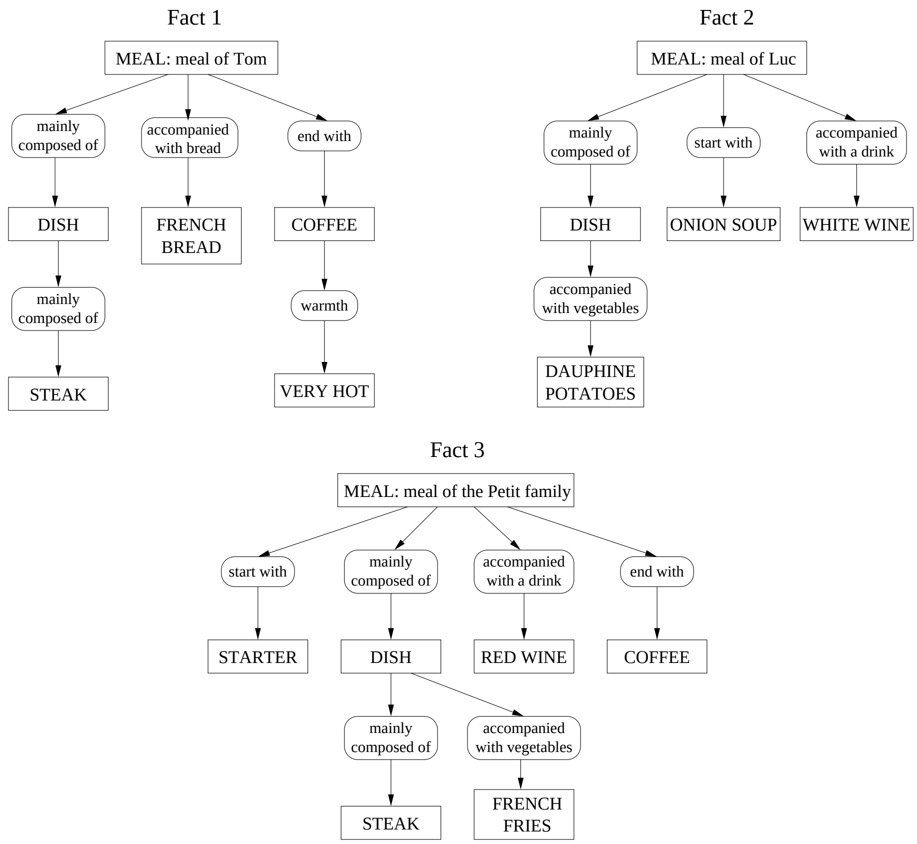

Fig. 3. Three facts corresponding to a French meal.

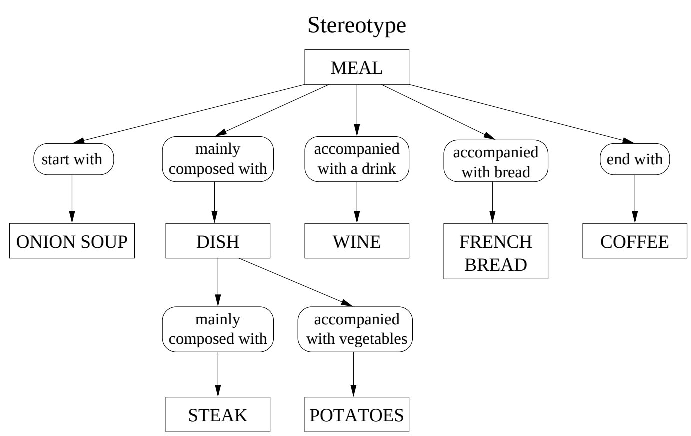

Fig. 4. A stereotype of a French meal.

the similarity between two conceptual graphs is proposed in [9]. However, in the present context, a measure which adds the differences is sufficient.

First let us recall the definition of *compatibility* given in [5]:

Definition 2. *Let conceptual graphs u_1 and u_2 have a common generalization v with projections $\pi_1 : v \rightarrow u_1$ and $\pi_2 : v \rightarrow u_2$. The two projections are said to be* compatible *if for each concept c in v, the following conditions are true:*

1. *$type(\pi_1 c) \cap type(\pi_2 c) > \perp$.*
2. *The referents of $\pi_1 c$ and $\pi_2 c$ conform to $type(\pi_1 c) \cap type(\pi_2 c)$.*
3. *If $referent(\pi_1 c)$ is the individual marker i, then $referent(\pi_2 c)$ is either i or $*$.*

We now consider that there is always only one least common generalization, i.e. only two projections that are compatible and maximally extended. It is easy to generalize with graphs having several least common generalizations.

We state the following theorem in order to link the notions of compatibility and default generalization:

Theorem 1. *Let conceptual graphs u_1 and u_2 have the least common generalization v with projections $\pi_1 : v \rightarrow u_1$ and $\pi_2 : v \rightarrow u_2$. π_1 and π_2 are compatible if and only if u_1 generalizes u_2 by default.*

Proof. If π_1 and π_2 are compatible then there exists a common specialization w of u_1 and u_2 (cf. theorem 3.5.7 [5]). According to definition 1, u_1 generalizes u_2 by default. Reciprocally, if u_1 generalizes u_2 by default then there exists a common specialization w. Suppose that π_1 and π_2 are not compatible. There therefore exists at least one concept in v with $type(\pi_1 c) \cap type(\pi_2 c) = \perp$, or with the referent of $\pi_1 c$ or $\pi_2 c$ not conform to $type(\pi_1 c) \cap type(\pi_2 c)$, or with $referent(\pi_1 c) = i$ and $referent(\pi_2 c) = j$, $i \neq j$. These three cases are absurd because they contradict the construction of w. Therefore, π_1 and π_2 are compatible.

Consider now the measure M_D counting the dissimilarities between two graphs u_1 and u_2. Let v be the least common generalization graph with projections $\pi_1 : v \rightarrow u_1$ and $\pi_2 : v \rightarrow u_2$. If π_1 and π_2 are not compatible then the measure $M_D(u_1, u_2)$ is fixed by convention with an infinite value noted M_∞ because one graph can't be generalized by default by the second one (cf. theorem 1). Otherwise $M_D(u_1, u_2)$ counts all the differences between the concepts and relations of u_1 and those of u_2. The measure is thus defined:

Definition 3. *Let conceptual graphs u_1 and u_2 have the least common generalization v with projections $\pi_1 : v \rightarrow u_1$ and $\pi_2 : v \rightarrow u_2$. The* measure of dissimilarities *$M_D(u_1, u_2)$ is equal to:*

1. *M_∞ if π_1 and π_2 are not compatible.*
2. *$C + T(u_1) + T(u_2)$ otherwise, where:*
   - *$C = |\{concept\ c \in v / type(\pi_1 c) \neq type(\pi_2 c)\ or\ referent(\pi_1 c) \neq referent(\pi_2 c)\}|$.*
   - *$T(u) = card(u) - card(v)$; $card(g)$ corresponds to the number of nodes (concepts and relations) of graph g.*

This measure presents the following properties:

Property 3. For any conceptual graph u, $M_D(u,u) = 0$.

Property 4. For any conceptuals graphs u and v, $M_D(u,v) = M_D(v,u)$.

2.2 Relative cover

We structure the memory into a set of stereotypes that summarizes the observations. Each fact, thanks to the measure defined above, is associated with the closest stereotype and completed with its description. *Relative cover* allows this calculation :

Definition 4. *The fact f is* covered *by the stereotype s* relative to *a set of stereotypes $\{s_1, ..., s_r\}$ if and only if:*

1. $\exists i$, $1 \leq i \leq r / s = s_i$.
2. $M_D(f,s) \neq M_\infty$.
3. $\forall k \neq i, 1 \leq k \leq r, M_D(f,s) < M_D(f,s_k)$.

Fig. 5 shows a set of three stereotypes that may be read s_1 = *The meal is accompanied by French bread with cheese and red wine from Bordeaux*, s_2 = *The meal is composed of steak with potatoes and beer* and s_3 = *The meal is composed of poultry with salad and mineral water*. Fig. 6 presents two facts to be classified within these stereotypes: f_1 = *The meal of Bob is composed of cheese and wine* and f_2 = *The meal of Léa is composed of salad and ends with a coffee without sugar*.

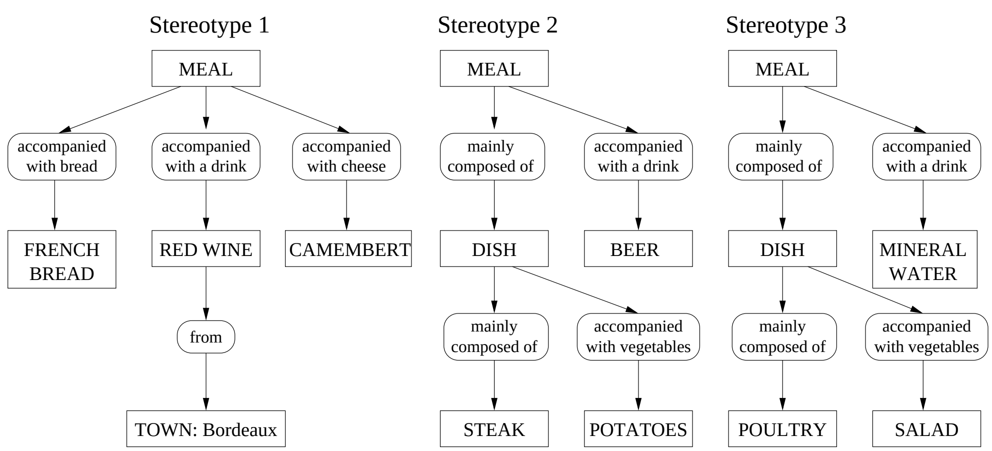

Fig. 5. Three stereotypes of a French meal.

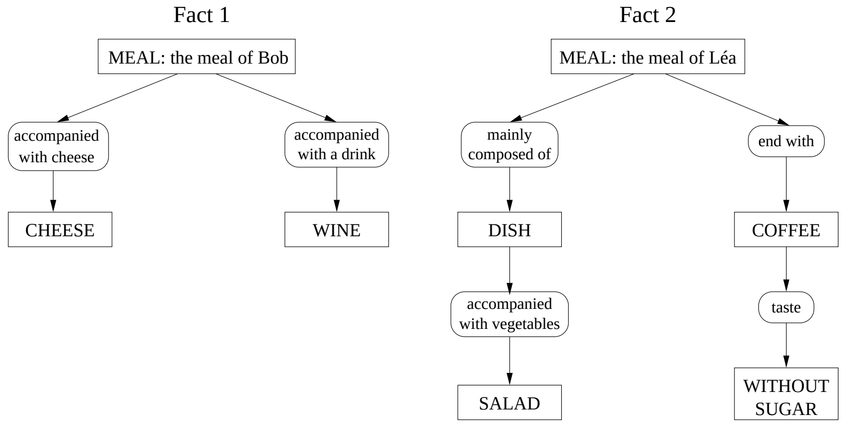

Fig. 6. Two facts to be classified.

Let us first consider the fact f_1 and the stereotype s_1. The common generalization v is *The meal is composed of cheese with wine*. $M_D(f_1, s_1)$ is therefore equal to $3 + (5 - 5) + (9 - 5) = 7$. The stereotype s_2 gives $M_D(f_1, s_2) = 2 + (5 - 3) + (9 - 3) = 10$. Finally, $M_D(f_1, s_3)$ is equal to M_∞ because *MINERAL WATER* $\cap$ *WINE* $= \perp$. f_1 is therefore covered by s_1 relative to s_1, s_2, s_3. This means that f_1 is associated with the cluster labeled with the stereotype s_1.

The same calculation on the second fact f_2 gives the measures $M_D(f_2, s_1) = 17$, $M_D(f_2, s_2) = M_\infty$ and $M_D(f_2, s_3) = 9$. The fact is therefore covered by s_3 relative to s_1, s_2, s_3 and classified in the cluster labeled with s_3. If no stereotype covers a fact relative to the others then the fact is classified in a "garbage" cluster.

This is the case if no stereotype generalizes it by default or if two stereotypes get the same best value for M_D, because one fact cannot be covered by two different stereotypes.

2.3 Search for the best set of stereotypes

The aim is to find the best set of stereotypes given a set of considered facts. To decide which set of conceptual graphs is the best, the definition of an evaluation function is needed.

The first function considered uses probabilities. It is very similar to the *Category Utility* measure (see [10]) which is used in the COBWEB system [3] to evaluate good partitions. But this measure is not really appropriate for sparse descriptions and uses the Attributes/Values formalism. Moreover, runtime cost was rather high.

What we propose here is a more classical way using the definition of an evaluation function based on the dissimilarities measure M_D. This function is minimized by the best set of stereotypes.

Definition 5. *Let F be the set of facts to be classified, S the set of stereotypes to be evaluated and s the function that associates the fact f of F to the stereotype $s(f)$ of S such that f is covered by $s(f)$ relative to S. The cost function h is defined below:*

$$h(S) = \sum_{f \in F} M_D(f, s(f))$$

There are several methods for exploring the search space. One is incremental and very similar to that used by Fisher in [3]. It starts from an empty set with no stereotypes, considers at each step a new fact to be covered and updates the set with some operators. There could be one operator which creates a new stereotype equal to the considered fact and another which modifies an existing stereotype to cover it. The "merge" operator is a little bit more difficult to implement, but the main difficulty lies in the "split" operator, especially when there are conceptual structures. The search in COBWEB is like a "hill-climbing" strategy and its robustness is largely due to these two specific operators. A more general approach is therefore certainly preferable.

The second option is to search for the best set of stereotypes using a "tabu" strategy. A neighborhood is calculated from a current solution with the assistance of permitted movements. These movements can be of low influence (add a concept to one stereotype, restrict a concept within another) and high influence (add or retract one stereotype of the set). The search uses short and long-term memory to avoid loops and to intelligently explore the search space.

3 Results

This section begins by examining the question of validation and goes on to present an application.

3.1 Validation

We have considered two ways to validate the proposed model.

The first one relies on real cases using knowledge from the domain studied. The experiments deal with a major issue which was of historic importance in France and so much has been written about it (cf. section 3.2). Articles from newspapers about this affair are very sparse and so fit perfectly into our cognitive model. We have to confirm that the stereotypes that have been built correspond well to the mental representations of the affair, as projected in the media. Preliminary results in the Attributes/Values formalism are promising and can be confronted to the knowledge of the domain provided by books and experts.

The second one gives a more formal and systematic method with the elaboration of artificial training sets. We begin with a set of "complete" facts, i.e. using all possible concepts and relations. Thus, a fact can have identical features with the others, compatible features, or even contradictory features. Each fact is then duplicated several times in order to form a set of perfectly coherent clusters of different sizes. Next, sparse descriptions are simulated by removing part of the data: certain facts are replaced randomly by more general ones. Finally, the clustering algorithm is performed on these new data in order to produce a set of stereotypes. The original considered facts were retrieved. We have also changed the proportion of missing values in order to test the robustness of this method.

Fig. 7 shows a graph which uses the Attributes/Values formalism to give the classification error rate with respect to the missing data rate:

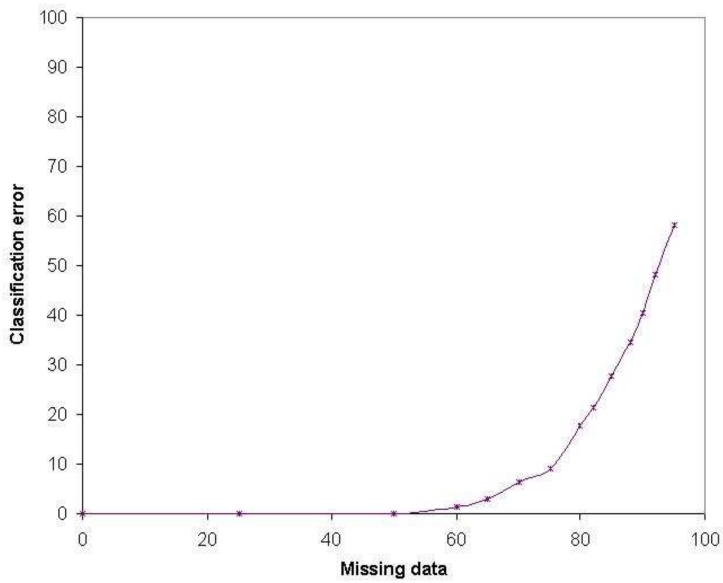

Fig. 7. Classification error using artificial training sets.

These experiments were carried out on three initial random facts until twenty runs in order to calculate averages. Each fact was duplicated fifty times. The classification error rate was counted according to the specific rate of removed descriptors. Errors are due to the facts which are not covered by the stereotype corresponding to the original fact they come from. These results prove the robustness of the method based on the concept of stereotypes, since the error rate is below 10% with up to 75% of missing data. That means that the three original facts were almost perfectly retrieved. Moreover, the rate does not exceed 28% with 85% of missing data, which is still satisfactory.

3.2 Applications

In today's information world there are many situations where data are so sparse that much interpretation is necessary.

The application we propose deals with an historic event, the famous miscarriage of justice known as the Dreyfys Affair which occurred at the end of the 19th century in France. In 1894 Captain Alfred Dreyfus, an officer on the French general staff, was accused of spying for Germany, France's opponent in the previous war. There were many articles about this very complex affair, bringing different views depending on the date, recent events, the newspaper political leanings. Thus, the liberal pro-Dreyfus *Le Siècle* expressed opinions which were diametrically opposed to those of the conservative anti-Dreyfus *L'éclair*. The facts we considered have been taken from these articles and translated into conceptual graphs, in order to build automatically a simplified model of the affair. The objective is to understand the influence of the press on the mental representations during this period.

Type hierarchies including 399 concepts and 174 relations were built for this specific context. In addition, a typical graph was proposed in order to translate the articles into facts. Fig. 8 shows an example of a graph. It represents an article from the newspaper *L'éclair* using the CoGITaNT library implemented by D. Genest and E. Salvat [11]:

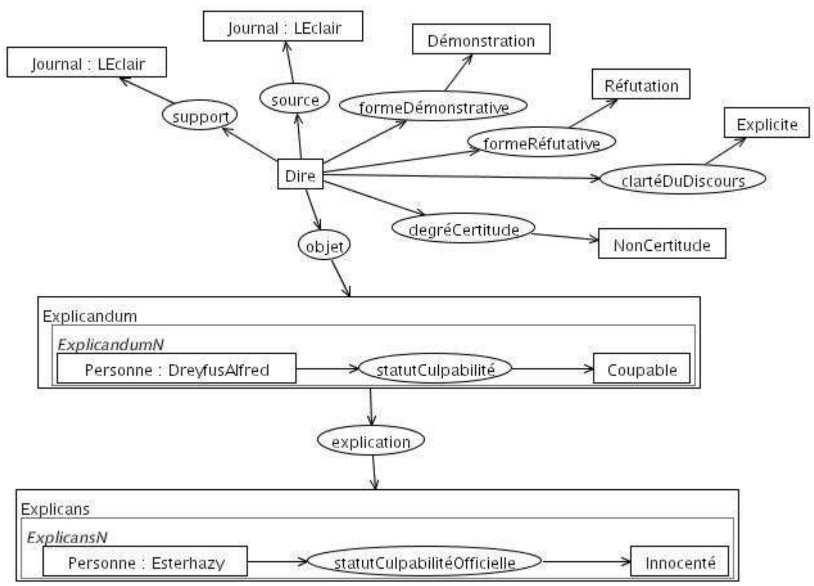

Fig. 8. A conceptual graph which translates a newspaper article.

It could be summarized as follows : *the article taken from the newspaper* L'éclair *explicitly asserts that Alfred Dreyfus is guilty because Esterhazy was proved innocent by the courts*. Once several articles have been translated in this way, stereotypes can be discovered using the methods proposed earlier.

4 Conclusion and future research

Flows of information play a key role in today's society. However, the value of information depends on its being interpreted correctly, and implicit knowledge has a considerable influence on this interpretation. This is particularly true of the media, such as newspapers, radio and television, where the information given is always incomplete. In this context we propose a cognitive model based on sets of stereotypes which summarize facts by "guessing" the missing values. Stereotypes are an alternative to prototypes and are more suitable for sparse descriptions. They rely on the notion of default generalization which relaxes constraints and makes possible the manipulation of such descriptions. Descriptions are then completed according to the closest stereotypes, with respect to the dissimilarities measure M_D. The present paper gives definitions of new operators which make this search possible using conceptual graphs formalism.

This work relates to the domain of social representations as introduced by Serge Moscovici in [12]. According to him, social representations are a sort of "common sense" knowledge which aims at inducing behaviors and allows communication between individuals. Social representations can be constructed with the help of sets of stereotypes, and the way these representations change can be studied through the media over different periods and social groups in comparison with such sets. This represents an unexplored way for enriching historical or social analysis.

Finally, this paper is closely related to the work of Herbert Simon [13]. Thus, the construction of internal representations is a fundamental problem in Artificial Intelligence, emerging from the research done on the notion of semantic in the 1960s. These sort of representations bring about desired actions from the large amounts of information gathered from the outside world. Stereotypes and default generalization can be used to summarize these sparse data for which more classical techniques are not appropriate.

Acknowledgments

The authors would particularly like to thank David Genest for his patience and useful advice concerning the CoGITaNT library, Rosalind Greenstein for reading and correcting the manuscript and Anne Remillieux for constructing the type hierarchies and graph examples.

References

1. Michalski, R.S.: Knowledge acquisition through conceptual clustering: A theoretical framework and algorithm for partitioning data into conjunctive concepts. In:

International Journal of Policy Analysis and Information Systems, number 4 (1980) 219–243.
2. Lebowitz, M.: Experiments with Incremental Concept Formation: UNIMEM. In: Machine Learning, number 2 (1987) 103–138.
3. Fisher, D.H.: Knowledge Acquisition Via Incremental Conceptual Clustering. In: Machine Learning, number 2 (1987) 139–172.
4. Reiter, R.: A logic for default reasoning. In: Artificial Intelligence, number 13 (1980) 81–132.
5. Sowa, J.F.: Conceptual Structures: Information Processing in Mind and Machine, Addison-Wesley Publishing Company (1984), Massachusetts, The Systems Programming Series.
6. Plotkin, G.D.: A Note on Inductive Generalization. In: Machine Intelligence, number 5. American Elsevier Publishing Company, Inc. (1970) 153–163. Meltzer, Bernard and Michie, Donald, New York.
7. Muggleton, S., De Raedt, L.: Inductive Logic Programming: Theory and Methods. In: Journal of Logic Programming, volume 19/20 (1994) 629–679.
8. Rosch, E.: Cognitive representations of semantic categories. In: Journal of Experimental Psychology: General, number 104 (1975) 192–232.
9. Zhong, J., Zhu H., Li J., Yu Y.: Conceptual Graph Matching for Semantic Search. In: Proceedings of the 10th International Conference on Conceptual Structures: Integration and Interfaces, Spring-Verlag (2002) 92–106.
10. Corter, J.E., Gluck, M.A.: Information, uncertainty, and the utility of categories, In: Proceedings of the Seventh Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates (1985) 283–287.
11. Genest, D., Salvat, E.: A platform allowing typed nested graphs: How cogito became cogitant. In: Proceedings of the Sixth International Conference on Conceptual Structures, Springer-Verlag (1998), Berlin.
12. Moscovici, S.: La psychanalyse: son image et son public. PUF (1961), Paris.
13. Simon, H.A., Siklóssy, L. (Eds.): Representation and meaning: experiments with information processing systems. Prentice-Hall, Inc. (1972). Englewood Cliffs, New Jersey.