



# Réseau sémantique universel... et logique floue<sup>1</sup>

© 1997 EPHITEQ

*Nous avons présenté l'année dernière (voir La Lettre n<sup>os</sup> 26 à 28) ce que pouvait être un véritable réseau sémantique, capable d'intégrer tous types de systèmes, et démontré qu'il était réalisable dans la pratique. Nous faisons ici le point sur l'avancement de la mise en œuvre, et ajoutons quelques réflexions sur des cas de figure rarement - en fait, pratiquement jamais - pris en compte dans un système à usage général.*

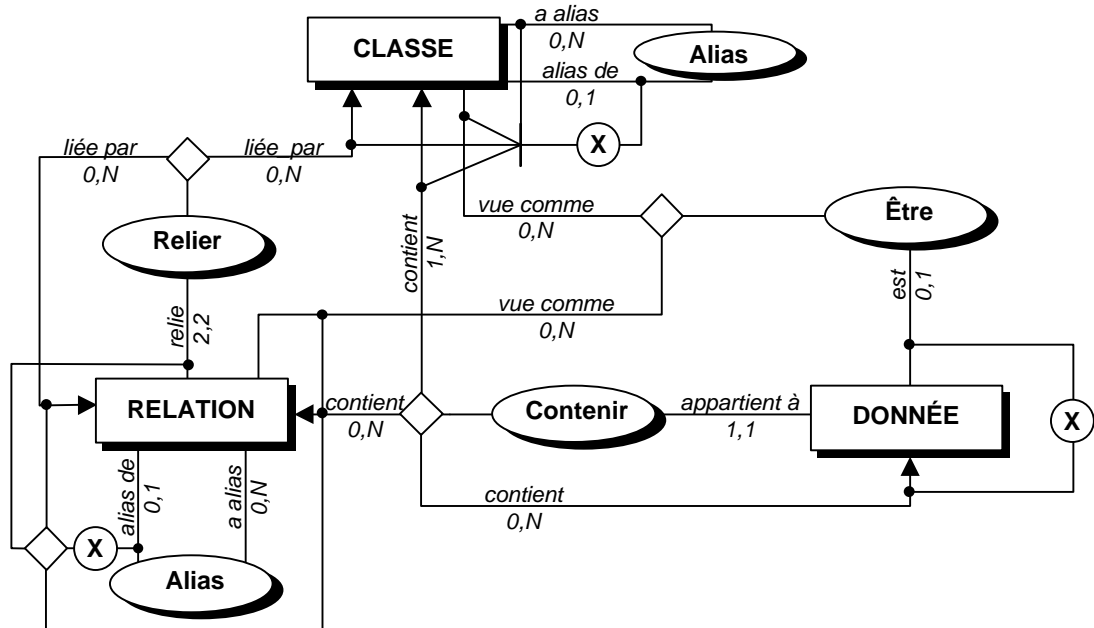
## Le point sur la mise en œuvre du prototype

À l'heure où vous lisez ces lignes, la phase 1 du prototype pratique de réseau sémantique universel est terminée, et la phase 2 vient de démarrer.

Avec l'achèvement de la phase 1, nous disposons d'un outil qui nous permet de créer, modifier, supprimer des entités, de les relier avec des relations, et de leur associer des attributs, sachant que chaque individu dispose de sa propre structure (jusqu'à 65000 champs par article, 1 milliard d'articles de chacune des trois classes de base), et que les attributs peuvent être arborescents (tableaux et agrégats). Tout cela tient dans trois tables relationnelles, les tables Entité et Relation ayant chacune quatre colonnes et deux index, la table Attributs ayant trois colonnes et un seul index. Il eût été possible de tout mettre dans une seule table de 4 colonnes et 2 index, mais il y aurait eu perte de place importante, tant pour les données que pour les index (du moins avec Paradox, le substrat choisi), et probablement également un impact sur les performances.

Cet outil nous permet d'effectuer un pré-chargement de la base avec les définitions indispensables des classes de base, sachant que les données sont définies "en dur" à ce niveau (le "Bios" de la base, en quelque sorte).

Pour mémoire, rappelons le MCD de cette mise en œuvre :



<sup>1</sup> Attention, il ne s'agit pas ici d'une présentation des principes de la logique floue. Mais pour ceux que cela intéresse, avis à ceux qui connaissent bien : quelques pages dans LA LETTRE sur ce sujet seront les bienvenues (surtout si l'auteur ne fait pas partie de la demi-douzaine d'adéliens qui ont fait l'intégralité des 4 numéros de 1997).

La phase 2 verra la mise en "DLL" du noyau du programme de la phase 1, l'intégralité des fonctions dépendantes du type de SGBD y étant intégré (donc, en cas de changement de SGBD support, ou de définition d'un support spécifique, seule cette DLL sera à modifier). Ce noyau sera utilisé par une application qui n'utilisera plus de définitions "en dur", mais chargera en mémoire les définitions stockées au niveau "bios" de la base, et permettra de les enrichir.

À la fin de cette phase, il sera possible de définir complètement un métamodèle tel que celui qui a été présenté dans La Lettre n°27.

La phase 3, convertissant le noyau de la phase 2 en une seconde DLL, permettra de créer un véritable outil de gestion de SGBD entité-relation et de créer des applications opérationnelles sans programmation - ou presque, puisqu'il ne s'agit toujours que d'un prototype -.

En pratique, la phase 2 sera doublée d'une phase 2bis, dans laquelle nous réaliserons une première application (donc avec des données définies "en dur") : **l'outil d'évaluation des AGL**, au sein de la commission ADELI du même nom.

Rendez-vous dans quelques mois...

## Quand le flou est de rigueur...

Nous avons déjà abordé la notion d'état flou pour une information (voir La Lettre n°26 , page 37). Nous allons ici développer le sujet.

Quand un numéro de téléphone n'est pas renseigné dans un article, est-ce parce que :

- On ne l'a pas fourni (c'est l'état par défaut "non renseigné") ?
- On n'a pas pu l'obtenir (état "inconnu") ? Et sait-on seulement si la personne a le téléphone ?
- La personne n'a pas le téléphone (état "sans objet") ?

Notre réseau sémantique universel permettra de différencier ces différentes causes d'absence. Mais il existe de nombreuses autres situations où on dispose d'informations, mais plus ou moins exactes, et plus ou moins fiable. Tentons d'en dresser une petite liste.

- On sait que l'information est comprise dans un certain ensemble de valeurs. Par exemple, on ne sait pas très bien si le véhicule est bleu ou vert, mais on est certain qu'il n'est ni rouge, ni jaune, ni blanc, ni gris...
- Inversement, on sait seulement que l'information n'est pas comprise dans un certain ensemble de valeurs. Par exemple, on ne se souvient pas de la couleur des cheveux du voleur, mais on est certain qu'il n'avait pas les cheveux blancs...
- On pense que l'information vaut tant, mais on n'en n'est pas du tout certain.
- On est presque certain que l'information a telle valeur...
- On connaît une partie de la valeur de l'information. Par exemple, on sait que le nom recherché commence par un B.
- On peut situer la valeur d'une information quantifiable dans certaines limites - ou en-dehors de ces limites - : inférieur à 100, mais supérieur à 10 et pas un multiple de 10...
- Pour une donnée logique, on se situe entre vrai et faux avec une certaine probabilité (ex : 90% de chances pour que ce soit vrai).
- On peut situer une information dans un ensemble de valeurs avec une répartition statistique (liste de valeurs avec pourcentage associé à chacune, ou fonction continue, genre courbe de Gauss...).
- On peut connaître une valeur avec une certaine marge d'erreur (ex : 120 à  $\pm 5\%$ ).
- etc.

Bref, il existe de nombreux cas autres que donnée renseignée et donnée manquante. Nous avons d'ores et déjà prévu de pouvoir intégrer ces différents cas en natif dans notre Réseau Sémantique Universel - pas dans le prototype en cours de réalisation, mais dans la version suivante, le principe de mise en œuvre étant déjà défini.

*Jean-Luc Blary*