

Lexique du cours “Knowledge Representation”

Extraction d'informations (IE) : C'est une branche de l'informatique qui s'intéresse à *extraire de l'information* structurée (par exemple sous forme de tables relationnelles) depuis des sources non-structurées (souvent du texte brut). L'intérêt de l'IE réside dans le fait que les traitements de données nécessitent souvent des informations structurées mais que les sources de données non structurées (par exemple le Web) sont nombreuses.

Exemple d'application de l'IE : récupération d'informations biographiques (naissance, mort, œuvre) depuis les articles de presse d'une personne célèbre.

Entité : Il s'agit de n'importe quel objet, personne, concept, *etc.*, du monde réel que l'on souhaite représenter. Par exemple, la *Joconde* ou *Simone de Beauvoir* sont des entités. La notion d'*écrivaine*, de *tableau* ou de *musée* peuvent aussi être des entités.

Dans notre formalisme, une entité n'a jamais de problème d'homonymie et chaque entité est censée avoir une sémantique précise et non ambiguë. Par exemple *Paris* a beaucoup d'homonymie (voir [https://fr.wikipedia.org/wiki/Paris_\(homonymie\)](https://fr.wikipedia.org/wiki/Paris_(homonymie))) mais chaque cas correspondrait à des entités différentes. Autre exemple, si l'on parle du Louvre, il peut y avoir plusieurs entités associées : une qui parle spécifiquement du bâtiment historique, une qui parle du musée, une qui parle de l'école, *etc.*.

Classe : Une classe est un type particulier d'entité, dont la signification est celle du trait commun d'une collection d'entités. Par exemple, *tableau*, *musée* ou *écrivaine* sont des classes.

Instance : Une entité E est une instance d'une classe C quand E est un cas particulier de C . Par exemple le *musée du Louvre* est une instance de *musée* et *La Joconde* est une instance de *tableau*. En revanche même si *La Joconde* est hébergée au *musée du Louvre*, elle n'en est pas une instance. On peut être instance de plusieurs classes, par exemple *La Joconde* est aussi instance de la classe *œuvres hébergées au Louvre* si cette classe existe.

Relation : Une relation R entre des classes C_1, \dots, C_k est un sous-ensemble de leur produit cartésien $C_1 \times \dots \times C_k$. Les classes C_1, \dots, C_k sont appelées le **domaine** de la relation tandis que le nombre k de classes est l'**arité** de la relation. Par exemple, on peut avoir la relation *est exposé dans* entre des tableaux et des musées qui indique où chaque tableau est exposé.

Relation binaire : Une relation binaire est une relation d'arité 2. Étant donnée une relation binaire entre C_1 et C_2 on note C_1 le **domaine** de la relation et C_2 son **image** (**range** en anglais).

Relation inverse : L'inverse d'une relation binaire, est la relation qui intervertit le range et le domaine. Par exemple l'inverse de la relation *<tableau> est exposé dans <musée>* sera *<musée> expose <tableau>*.

Relation fonctionnelle et inverse fonctionnelle : Une relation est fonctionnelle si, à tout élément du domaine, il existe au plus un élément de l'image qui lui est associé. Une relation est inverse fonctionnelle si sa relation inverse est fonctionnelle. Par exemple, *est exposé dans* est une relation fonctionnelle (car les tableaux ne peuvent pas être à plusieurs endroits).

Graphe de connaissance (Knowledge Base ou KB en anglais) : Une KB est la donnée d'un ensemble de relations sur des entités. Il peut se représenter comme un graphe orienté où les arcs ont des labels, ou alors comme un ensemble de triplets.

Schéma : Une KB peut avoir un schéma qui spécifie pour chaque relation la classe des éléments du domaine et du range.

Canonicité : La canonicité d'une entité ou d'une relation dans une KB est la propriété qu'il n'existe qu'elle dans la KB qui représente la même chose / personne / concept / relation. Par exemple si ma KB contient deux entités *De Beauvoir* et *Simone de Beauvoir* qui traitent toutes deux de l'écrivaine, ma KB n'est pas canonique. De même si ma KB contient les relations *est exposé dans* et *expose l'œuvre* alors elle n'est pas canonique.