

TP4

Acquisition de connaissances

I Le système solaire

Question 2 : Si l'on intègre Pluton parmi les planètes, elle devient l'unique représentante des planètes petites, éloignées et possédant des satellites. Ainsi le treillis change et se complexifie, afin de s'adapter à ce nouvel élément.

Dans la FCA, l'exception à plus de poids car on crée un nouveau concept. En règle d'association nous cherchons les tendances, ainsi une exception comme Pluton ne générera pas forcément de nouvelles règles.

Question 3 : En mode "show only exact matches", Siena affiche les objets dont le système est entièrement défini (qui possède tous les attributs).

En mode "show all matches", Siena affiche également les objets appartenant aux combinaisons d'attributs intermédiaires (non complètes).

II Echelle conceptuelle

Question 4 : Taille des disques durs : échelle ordinale

Types de bus système : nominale

Moyen de distribution : booléenne (2 attributs à 1 ou 0)

III Recensement Américain

Question 5 : Il existe trois formes équivalentes pour représenter les données : le treillis, la base d'implication et le tableau de données.

Les règles d'association contiennent des règles redondantes n'apportant pas plus d'informations. La base d'implication représente un sous-ensemble des règles d'associations. Enfin, la base minimale permet de reconstruire toutes les données.

Question 6 : Le treillis est trop grand pour être exploitable. Un treillis de ce genre peut permettre de trouver les implications exactes. L'analyse de concept permet de créer de nouveaux treillis complets et non redondant (à l'image de la base d'implication Duquenne-guiges) grâce à des algorithmes comme le Next-Closure.

IV Base de comics avec Camelis

Question 7 : A l'ouverture on voit que chaque objet est décrit par son dessinateur, son éditeur, son genre, son nom de série, sa note sur 5 et son scénariste.

Question 8 : La base est composée de 627 objets (le premier étant une description), de 1699 attributs et sa taille est de 43950 paires de type (object,feature) (10021 mots). Le nombre moyen d'attributs par objet est de 69.

Question 9 : Pour trouver le nombre de séries par éditeur, il suffit de regarder dans l'extension de l'éditeur. Ainsi, Akileos a 12 séries, Albin Michel en a 11, Bamboo 11, Bethy 11 etc...

Le dessinateur ayant le plus de Comics dans ce contexte est Richard Corben.

Question 10 : Pour obtenir les comics ayant la meilleure note (qui est 4 dans ce contexte) on effectue la requête suivante : ('Note /5' is "4 (2 avis)" or 'Note /5' is "4 (1 avis)").

('Note /5' is "4 (1 avis)" or 'Note /5' is "4 (2 avis)")

Appliquer

The screenshot shows the Camelis software interface. At the top, there is a search bar with the query 'Dessinateur ?' and a list of results. Below the search bar, there are two tabs: 'Picto (0)' and 'Texte (89)'. The 'Texte (89)' tab is selected, showing a list of results for the query. The results are displayed in a table with columns for the number of results, the number of objects, and the text description. The results are sorted by the number of objects in descending order.

all	627	89	Dessinateur ?
			<bottom>
4	3		Dessinateur is "Caniff (Milton)"
6	2		Dessinateur is "Clowes (Daniel)"
13	1		Dessinateur is "Collectif"
15	3		Dessinateur is "Corben (Richard)"
8	1		Dessinateur is "Crumb (Robert)"
13	1		Dessinateur is "Eisner (Will)"
6	1		Dessinateur is "Gibbons (Dave)"
6	3		Dessinateur is "Kirby (Jack)"
3	2		Dessinateur is "Kordey (Igor)"
5	1		Dessinateur is "Mazzucchelli (David)"
11	1		Dessinateur is "Miller (Frank)"
2	2		Dessinateur is "Quitely (Frank)"
3	2		Dessinateur is "Raymond (Alex)"
5	1		Dessinateur is "Sacco (Joe)"
7	1		Dessinateur is "Sale (Tim)"
			<down>
627	89		Editeur ?
12	1		Editeur is "Akileos"
26	3		Editeur is "Comics USA"
13	1		Editeur is "Cà et Là"
72	3		Editeur is "Delcourt"
20	2		Editeur is "Editions USA"

On en déduit que les dessinateurs des Comics les plus appréciés sont Milton Caniff, Daniel Clowes, Collectif, Richard Corben. Leurs scénaristes sont Kurt Busiek, Milton Cannif, Daniel Clowes et Collectif.

Question 11 : Les série contenant "Batman" dans leurs titres et n'étant pas éditées ni par Paninin Comics ni Semic peuvent être trouvées grâce à la requête : 'Nom série' contains "Batman" and not Editeur is "Semic" and not Editeur is "Panini Comics". On obtient alors : Batman - Année 1.

Question 12 : Pour trouver les séries n'ayant pas de genre, on sélectionne tous les genres possibles en faisant un "ou" entre chaque, puis on ajoute un not devant. On obtient 16 résultats.

Question 13 :

('Nom série' contains "blanche" or 'Nom série' contains "noire" or 'Nom série' contains "mort" or 'Nom série' contains "vie") and not 'Nom série' contains "une" and not 'Nom série' contains "la"

☐ NOT ☒ = ☐ >= ☐ <= Zoom Pivot Picto (0) Texto (7)

all 7 ▸ Dessinateur ?
7 ▸ Editeur ?
7 ▸ Genre ?
7 ▸ 'Nom série' ?
7 ▸ 'Note /5' ?
7 ▸ 'Scénariste' ?
7 ▸ source ?

< < Results: 1 - 7 / 7

A l'ombre des tours mortes
Mon dernier jour au Vietnam
Mort@17
Parlez-moi de Mort
Quand je serai vieille et autres histoires
Terreurs Noires
Veuve noire

Question 14 : On effectue la requête 'Nom série' is « Batman – Dark Knight » puis on double-clique sur le résultat pour obtenir l'ensemble de ses propriétés :

source path is "/home/maud/Documents/TP4" and source ext is "csv" and source name is "catalogue_Comics" and source dir is "TP4" and source dir is "Documents" and source dir is "maud" and source dir is "home" and source size = 62513 and source last_modif date = 11 mar 2015 and source last_status_change date = 11 mar 2015 and source last_access date = 11 mar 2015 and source user is "maud" and source group is "maud" and source perm = rw - - and source part and 'Nom série' is "Batman - Dark Knight" and 'Note /5' is "4.45 (22 avis)" and Editeur is "Delcourt" and Genre is "Policier/Thriller" and 'Scénariste' is "Miller (Frank)" and Dessinateur is "Miller (Frank)"

On peut ensuite effectuer des requêtes en sélectionnant un certain nombre de ces propriétés afin de trouver les comics les partageant, qui seront alors considérés comme proches.

Question 15 :

SQL, Hadoop, Spark, Microsoft Access, Neo4j...