

Acquisition de Connaissances TP4 Analyse de Concepts

II - Système Solaire

Question 2 :

Files

Update: Clear dependent

Document

planetes.txt
Context

	A	B	C	D	E	F	G	H
		petite	moyenne	grande	procheS...	loinSoleil	satellite	noSatellite
Mercure		X			X			X
Venus		X			X			X
Terre		X			X		X	
Mars		X			X		X	
Jupiter				X		X	X	
Saturne				X		X	X	
Uranus			X			X	X	
Neptune			X			X	X	
Pluton		X				X	X	

Parameter Value

Show arrow rel... don't show

Compressed

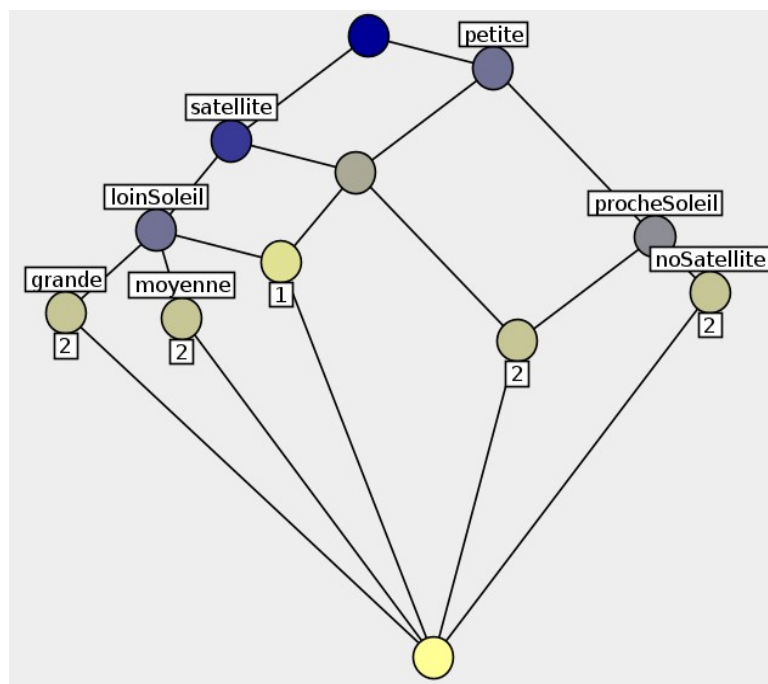
Object count 9

Attribute count 7

Si l'on considère Pluton comme une planète, il devient l'unique représentant des planètes éloignées du Soleil, avec satellite et petites.

On observe que la FCA génère toutes les combinaisons d'attributs ayant au moins un représentant parmi les objets.

Question 3 :



En mode "show only exact matches", Siena affiche les objets seulement lorsque toutes les combinaisons d'attributs ont été effectués et donc que le système est entièrement défini.

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

III - Echelle conceptuelle

Question 4 :

Taille des disques durs : numérique (discrétisés)
Types de bus système : nominal
Moyen de distribution : booléen (2 attributs à 1 ou 0)

IV - Recensement américain : Implications-Règles d'association

Question 5 :

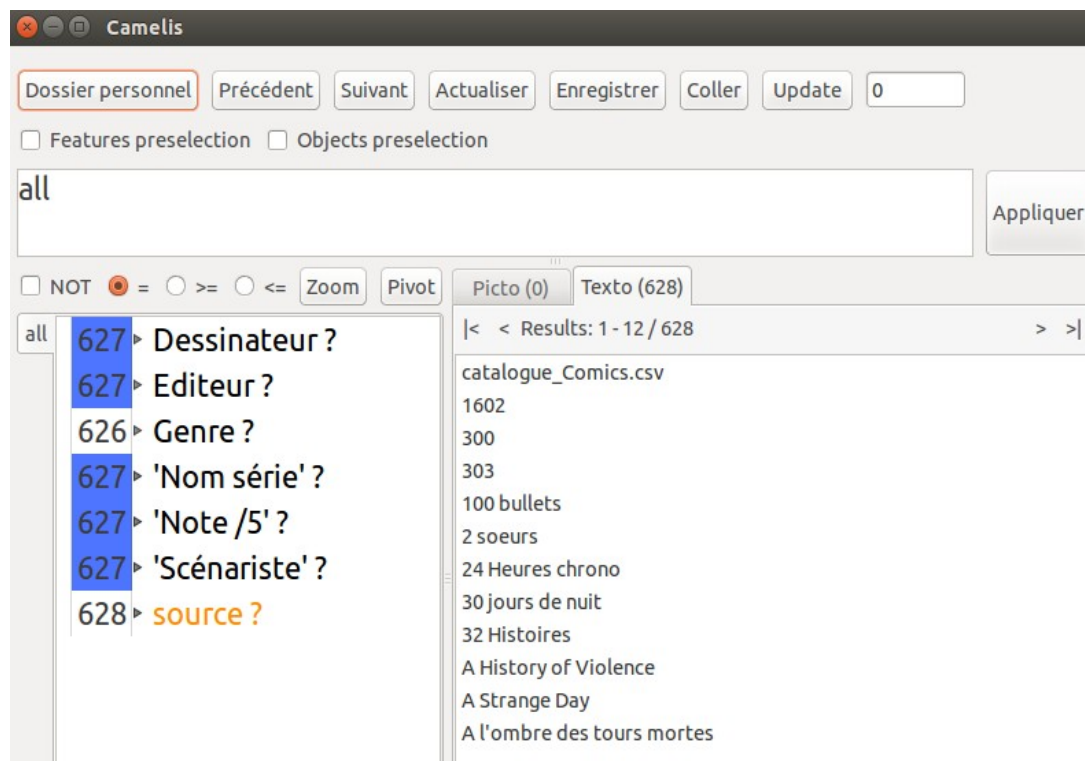
On constate que les règles extraites et la base d'implication du fichier de recensement présentent des formes très proches, mais que les règles rendent compte de plus de leurs précisions sous forme statistique.

Question 6 :

Le treillis est beaucoup trop grand pour être exploitable visuellement, ses dimensions le rendent illisibles. Un tel treillis peut permettre de trouver les implications exactes. L'analyse de concept met à disposition pour l'analyse de tels treillis des algorithmes, tels que le Next-Closure, qui permettent de créer une nouvelle représentation du treillis, complète et non redondante (à l'image de la base d'implication Duquenne-guiges).

V - Base de comics gérée avec Camelis

Question 7 :



Dès l'ouverture du fichier, on constate que certaines informations ont déjà été triées : le logiciel a séparé et classé par ordre alphabétique les différentes données (dessinateur, éditeur, etc.). Pour le moment aucune requête n'a été appliquée, donc tous les comics apparaissent.

Question 8 :

Notre contexte courant comporte 628 objets (objects), 1699 attributs (features), 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 le nombre de séries par éditeur, il suffit de cliquer sur « Editeur ? » dans la liste de gauche. On obtient alors la liste des éditeurs, et à côté de chaque, le nombre de séries qu'ils éditent. Par exemple, Akileos édite 12 comics, Albin Michel 11, et ainsi de suite. Les résultats sont visibles ci-après :

627	▸ Dessinateur ?	< < Results: 1 - 12 / 628
627	▾ Editeur ?	catalogue_Comics.csv
	<bottom>	1602
12	Editeur is "Akileos"	300
11	Editeur is "Albin Michel"	303
11	▸ Editeur is "Bamboo"	100 bullets
11	▸ Editeur is "Bethy"	2 soeurs
26	▸ Editeur is "Comics USA"	24 Heures chrono
11	▸ Editeur is "Cornélius"	30 jours de nuit
13	▸ Editeur is "Cà et Là"	32 Histoires
72	▸ Editeur is "Delcourt"	A History of Violence
20	▸ Editeur is "Editions USA"	A Strange Day
		A l'ombre des tours mortes

Pour connaître le nombre de comics par dessinateur, on utilise la même méthode en cliquant sur « Dessinateur ? » et on trouve alors que Corben (Richard) est celui ayant le plus de comics (15 pour être précis) dans ce contexte.

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)"). On obtient ensuite les dessinateurs et les scénaristes en cliquant respectivement sur « Dessinateur ? » et « Scénariste ? ». On voit alors que les dessinateurs ayant dessiné le plus de comics à succès (3) sont Caniff (Milton), Corben (Richard) et Kirby (Jack). Les scénaristes à succès sont Caniff (Milton), Elis (Warren) et Moore (Alan).

('Note /5' is "4 (2 avis)" or 'Note /5' is "4 (1 avis)")		
<input type="checkbox"/> NOT <input checked="" type="radio"/> = <input type="radio"/> >= <input type="radio"/> <= Zoom Pivot Picto (0) Texto (89)		
all	89 ▸ Dessinateur ?	< < Results: 1 - 12 / 89
	89 ▸ Editeur ?	2 soeurs
	88 ▸ Genre ?	32 Histoires
	89 ▸ 'Nom série' ?	Aliens versus Predator
	89 ▾ 'Note /5' ?	Angel fire
	72 ▸ 'Note /5' is "4 (1 avis)"	Arrowsmith
	17 'Note /5' is "4 (2 avis)"	Astro City
	89 ▸ 'Scénariste' ?	Au Coeur de l'Empire - L'Héritage de Luther
	89 ▸ source ?	Authority
		Aventures de Luther Arkwright (les)
		Batman - Proie
		Batman - Secrets
		Batman!

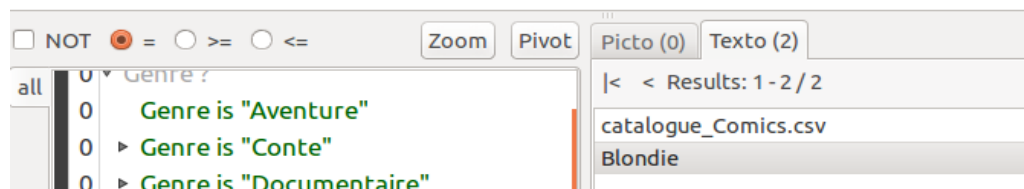
Question 11 :

On obtient la réponse à cette question par le biais de la requête : 'Nom série' contains "Batman" and not Editeur is "Semic" and not Editeur is "Panini Comics". On obtient 16 résultats dont les premiers sont montrés sur la capture d'écran ci-dessous :



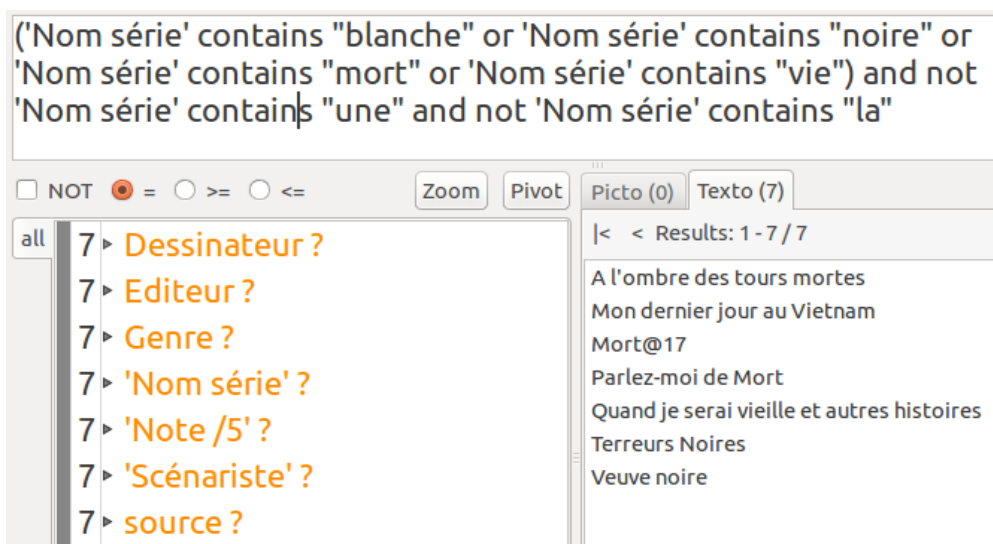
Question 12 :

not Genre is "Aventure" and not Genre is "Conte" and not Genre is "Documentaire" and not Genre is "Fantastique" and not Genre is "Fantasy" and not Genre is "Historique" and not Genre is "Humour" and not Genre is "Inclassable" and not Genre is "Jeunesse" and not Genre is "Médiéval Fantastique" and not Genre is "Policier/Thriller" and not Genre is "Roman Graphique" and not Genre is "Science-Fiction" and not Genre is "Western"



On pense qu'il existe une requête plus simple pour répondre à cette question mais nous n'en avons pas connaissance.

Question 13 :



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...