



HAL
open science

Expression de requêtes en graphes conceptuels à partir de mots-clés et de patrons

Catherine Comparot, Ollivier Haemmerlé, Nathalie Jane Hernandez

► **To cite this version:**

Catherine Comparot, Ollivier Haemmerlé, Nathalie Jane Hernandez. Expression de requêtes en graphes conceptuels à partir de mots-clés et de patrons. IC - 21èmes Journées francophones d'Ingénierie des Connaissances, Jun 2010, Nîmes, France. hal-01022623

HAL Id: hal-01022623

<https://inria.hal.science/hal-01022623v1>

Submitted on 10 Jul 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Expression de requêtes en graphes conceptuels à partir de mots-clés et de patrons

Cathy Comparot
Olivier Haemmerlé
Nathalie Hernandez

IRIT – équipe IC3
Université Toulouse 2
Département Maths-Info
5 allées Antonio Machado
31058 TOULOUSE Cedex

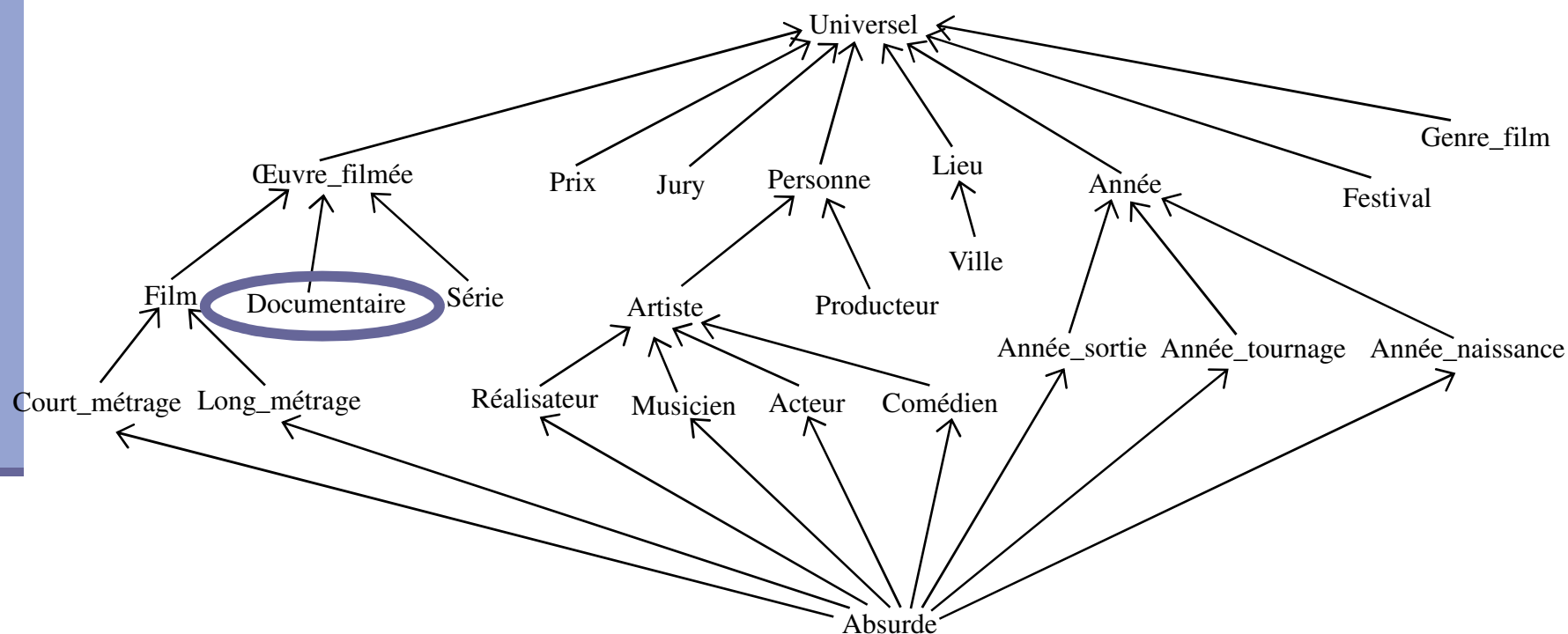


Préambule : le modèle des graphes conceptuels

- Deux ensembles distincts de connaissances
 - le support, ou connaissances terminologiques (ou ontologie)
 - vocabulaire
 - contraintes d'utilisation du vocabulaire
 - les graphes, ou connaissances assertionnelles
 - faits proprement dits

GC – Le support $S = \{Tc, Tr, M\}$

- Tc est l'ensemble des types de concepts



termes associés : « documentaire », « film documentaire », « reportage »

GC – Le support (2)

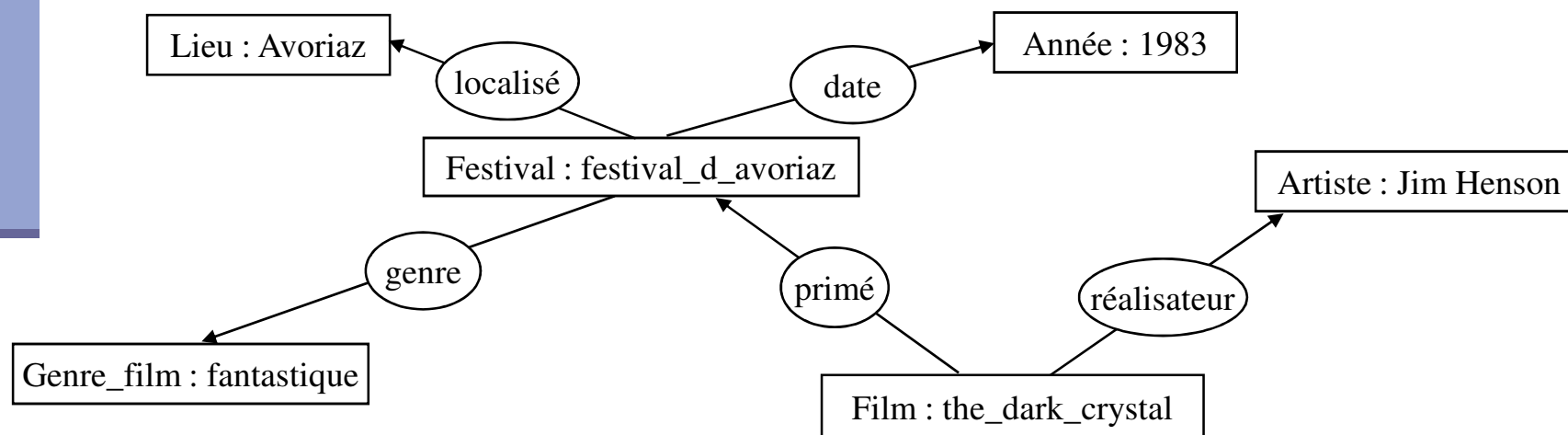
- Tr est l'ensemble des types de relations
 - Description des liens entre concepts
 - Partiellement ordonnés
 - genre (Festival, Genre_film)
 - localisé (Festival, Lieu)
 - reçoit (Artiste, Prix)

GC – Le support (3)

- M est l'ensemble des marqueurs
 - Les individus (ou instances) d'un concept donné
 - « Jim Henson » (instance du type de concept « Artiste »)
 - « The Dark Crystal » (instance de « Film »)
 - « Festival de Cannes » (instance de « Festival »)
 - Le marqueur générique permet de désigner un indéterminé
 - noté * ou omis
 - *un* Film, *un* Acteur...

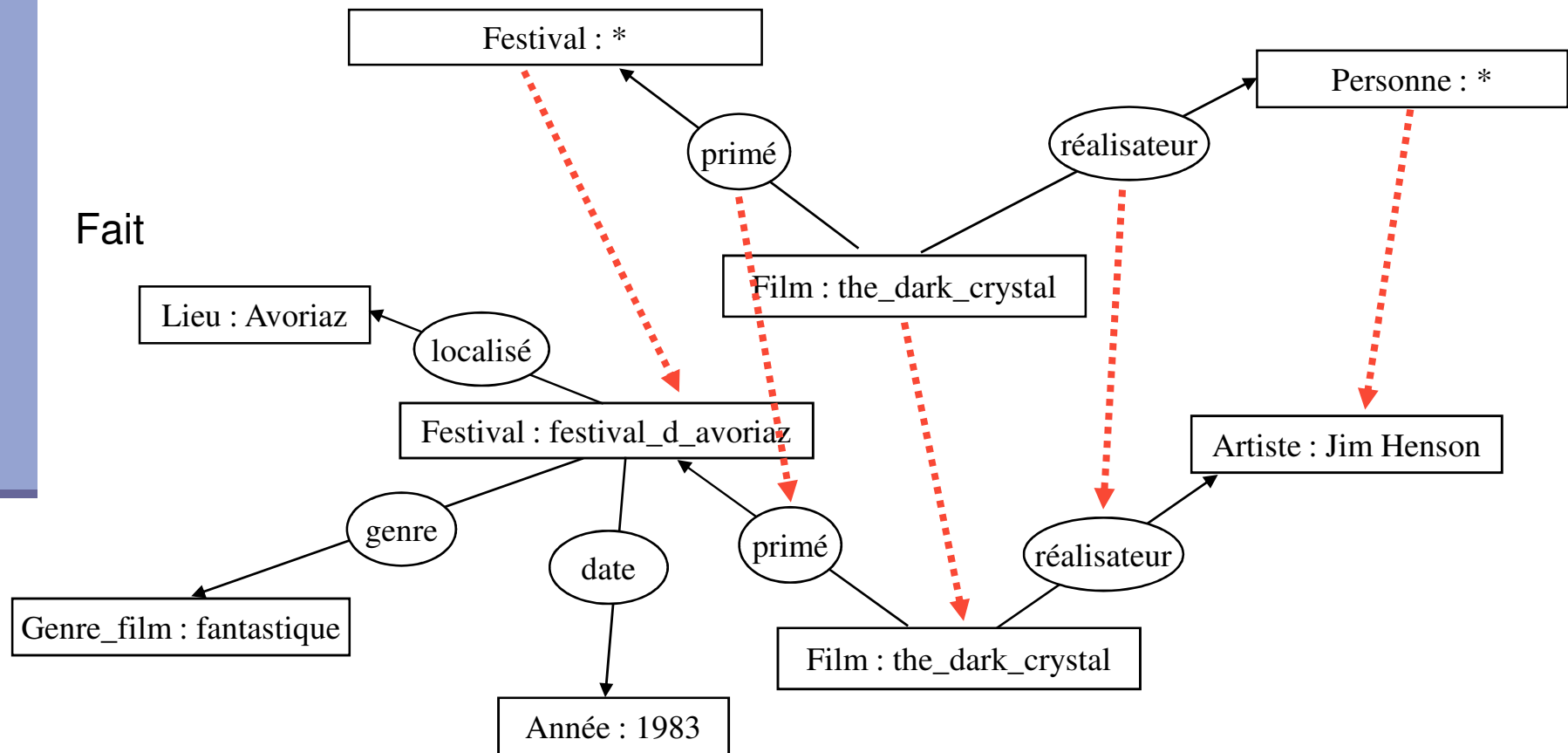
termes associés : « Festival de Cannes », « Cannes », « Festival international du film »

Les graphes conceptuels



Projection

Requête



Objectif général



Auriez-vous l'extrême amabilité de me dire dans quel film Patrick Dewaere s'extasie sur la mélodie hongroise de Schubert ???

```
PREFIX rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
PREFIX foaf: http://xmlns.com/foaf/0.1/
PREFIX movie: http://...
PREFIX dbpedia: http://...
PREFIX dc: http://purl.org/dc/elements/1.1/
SELECT DISTINCT ?titre
WHERE {
    ?personne rdf:type foaf:Person .
    ?personne foaf:name foaf:Patrick_Deware.
    ?music dbpedia:composer dbpedia:Schubert.
    ?film rdf:type movie:Film .
    ?film movie:actor ?personne.
    ?film movie:titre ?titre .
}
```

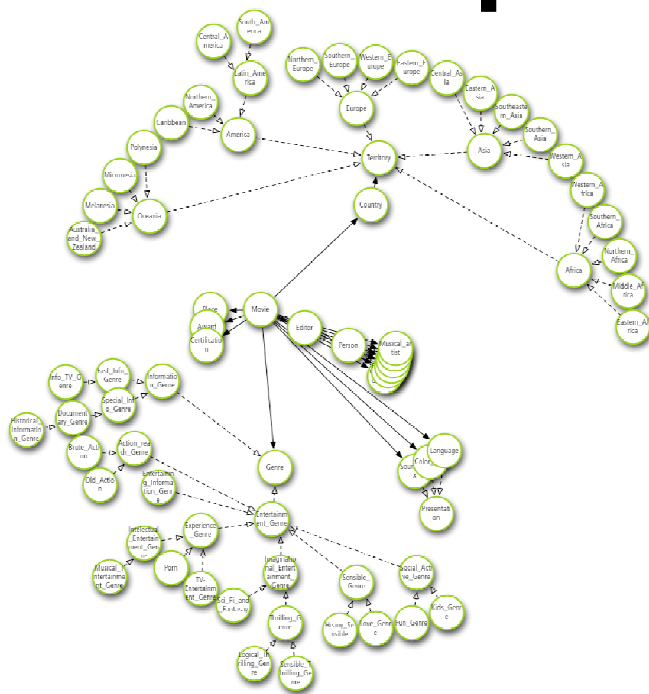
Préparez vos mouchoirs !!!

Objectif dégradé (1^{ère} étape)

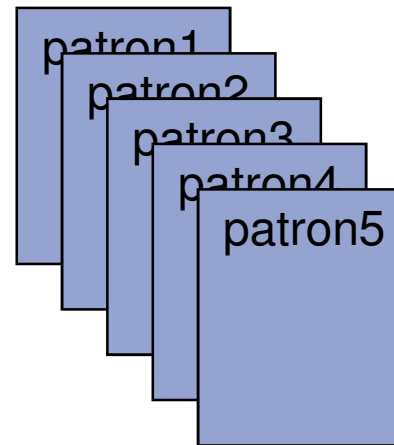
LN → mots-clés

Dewaere « mélodie hongroise » Schubert

+

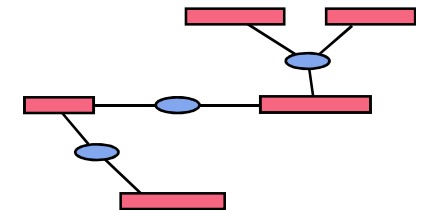


+

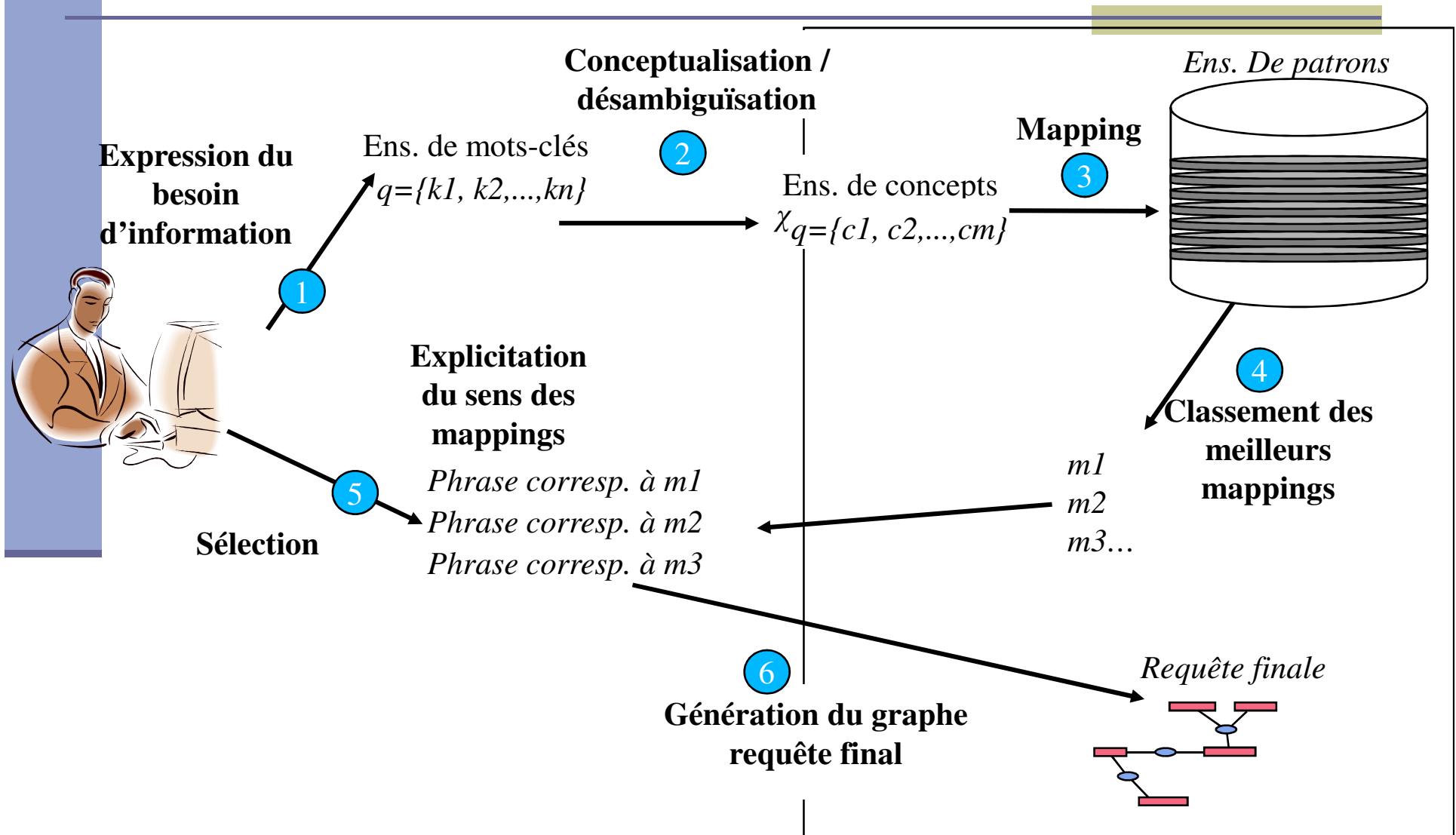


=

Requête GC



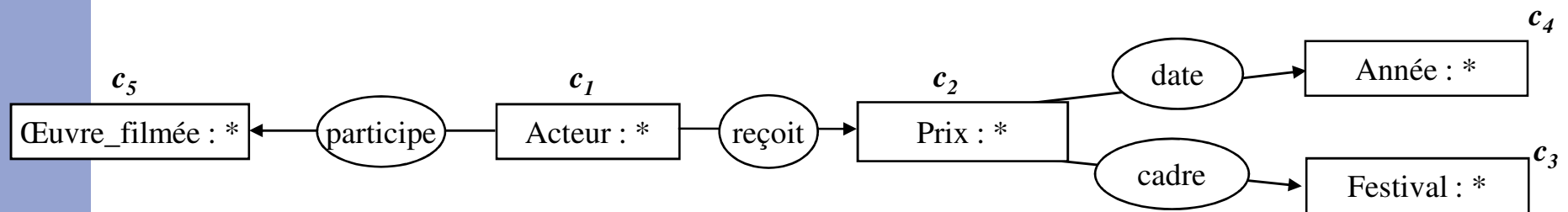
Proposition actuelle



Qu'appelle-t-on patron ? (1)



Qu'appelle-t-on patron ? (2)

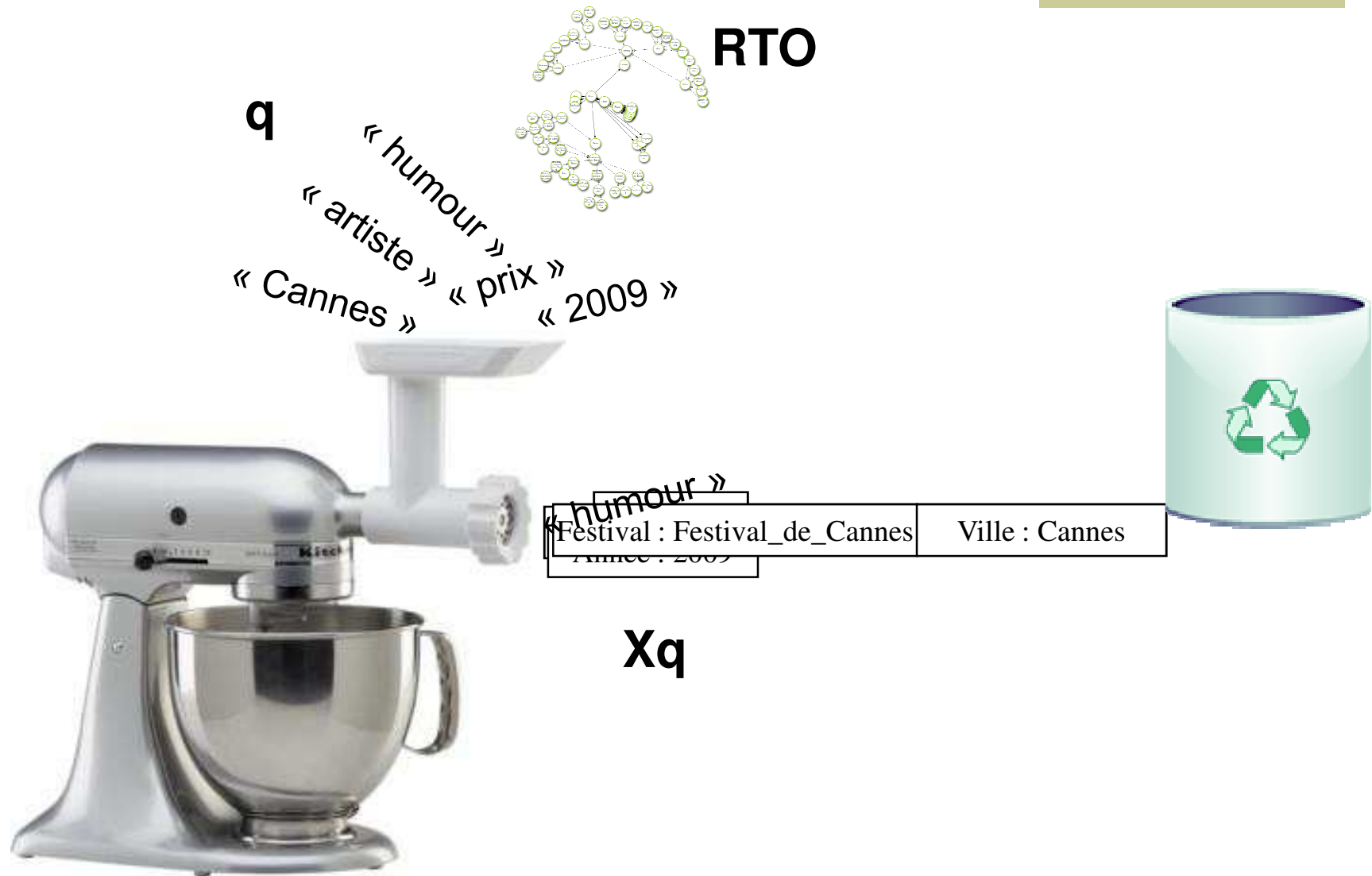


- $p = \{G_p, C_p, S_p\}$
- $C_p = \{c_1, c_2, c_3, c_4, c_5\}$
- $S = \ll [\text{Un acteur}]_{w1} \text{ ayant obtenu } [\text{un prix}]_{w2} \text{ lors de } [\text{un festival}]_{w3} \text{ en } [\text{une année}]_{w4} \text{ pour } [\text{une œuvre filmée}]_{w5} \gg$

Etape 1 : expression de la requête utilisateur

- $q = \{\ll \text{Cannes} \gg, \ll \text{prix} \gg, \ll \text{artiste} \gg, \ll \text{2009} \gg, \ll \text{humour} \gg\}$
- L'utilisateur souhaite obtenir des informations sur les prix obtenus par des humoristes lors du festival de Cannes 2009

Etape 2 : conceptualisation et désambiguïsation de la requête



Etape 3 : matching requête / patrons

- Mise en correspondance de la requête conceptualisée avec les concepts qualifiants des patrons (c-mapping)
 - sommets concepts de la requête conceptualisée partitionnés en 4
 - sommets identiques à un sommet du patron
 - sommets spécialisant un sommet du patron
 - sommets généralisant un sommet du patron
 - sommets non mis en correspondance

Etape 3 (suite)

- Identiques requête/patron
 - [Prix : *] → [Prix : *]
- Requête < patron
 - [Année : 2009] → [Année : *]
 - [Festival : festival_de_cannes] → [Festival : *]
- Requête > patron
 - [Artiste : *] → [Acteur : *]

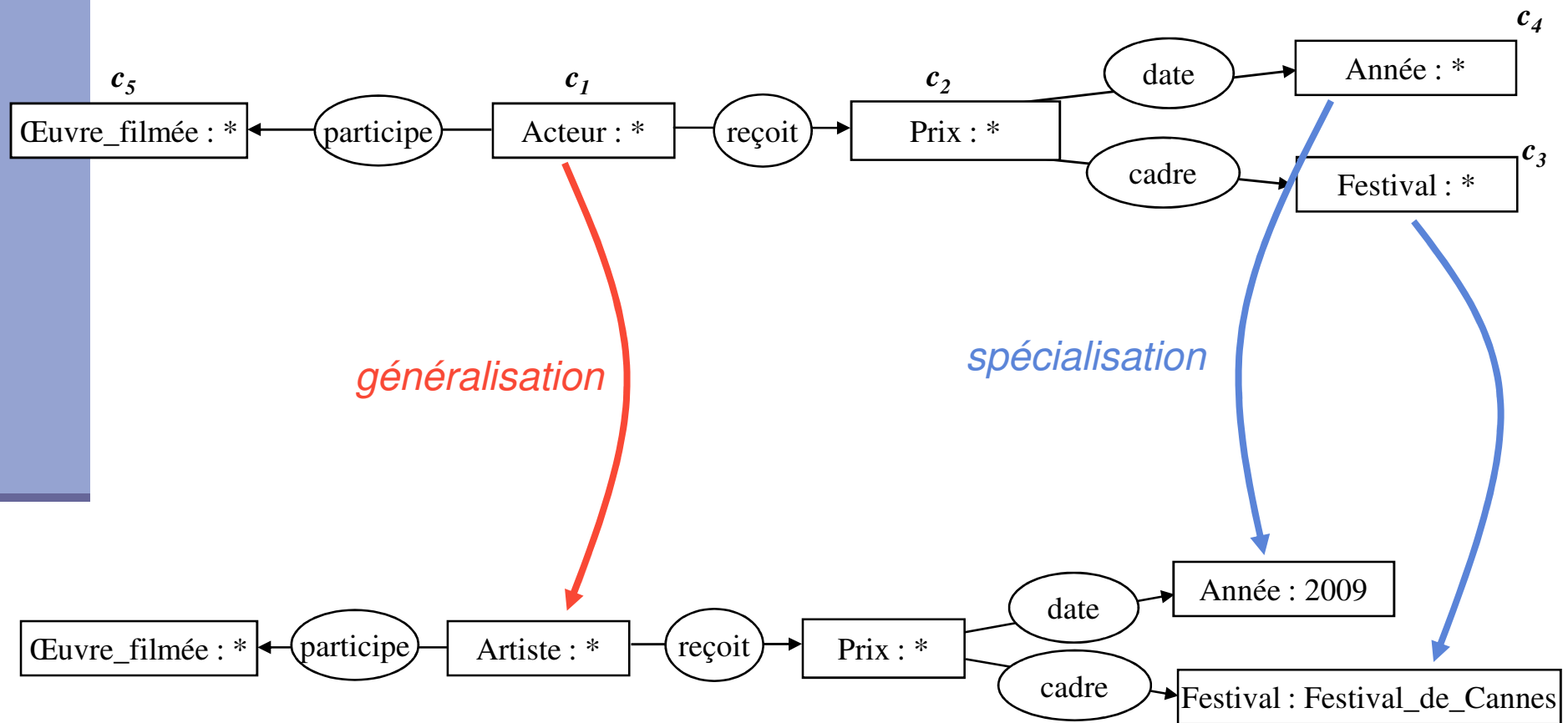
Etape 4 : classement des c-mappings

- Classement des différents c-mappings selon une mesure de pertinence par rapport à la requête conceptualisée
 - $|C_{q=}|$ ↑ pertinence du c-mapping ↑
 - $|C_{q<}|$ ↑ pertinence du c-mapping ↑
 - spécialisation du patron conformément à la requête produit un graphe requête pertinent
 - $|C_{q>}|$ ↑ pertinence du c-mapping ↑
 - généralisation du patron conformément à la requête peut produire un graphe requête non pertinent
 - $|p_{only}|$ ↑ pertinence du c-mapping ↓
 - plus important est l'ensemble des concepts non retrouvés dans p , moins pertinent est le mapping

Etape 5 : génération des phrases décrivant les requêtes

- Pour chacun des c-mappings les mieux classés
 - génération de la phrase expliquant la requête
- 1. « ***Un artiste*** ayant obtenu [un prix] lors de ***Festival de Cannes*** en ***2009*** pour [une œuvre filmée] ». « humour » ignoré
- 2. « ***Festival de Cannes*** de [un genre] ayant lieu à [un lieu] en ***2009*** dont [un artiste] fait partie de [un jury] ». « humour » ignoré



Etape 6 : construction de la requête GC



Exécution de la requête

- Opération de projection
- Graphe acyclique dans graphe quelconque

Evaluation

- 160 requêtes d'utilisateurs lambda
 - mots clés / description en langue nat.
- Support conçus à partir de  + 
- 1/3 de requêtes analysées, 13 patrons construits manuellement
- Familles de requêtes
 - Les 2/3 restants sont couverts à 73% par les 13 patrons (27 % restants +/- caractérisés)
- Pertinence des requêtes générées
 - Comparaison des GC générés avec les GC construits manuellement à partir de la description en langue nat.
 - Pour 88% des requêtes, la « bonne » requête apparaît dans les 3 premières propositions

Conclusion

- Premier prototype en GC
 - 2^{ème} proto SPARQL
- Tests en cours en SPARQL sur DBpedia
- Résultats expérimentaux encourageants
 - les requêtes fréquemment posées peuvent servir de base à la génération de nouvelles requêtes
 - le classement des requêtes générées est satisfaisant



Travail en cours ou court terme

- Chasser les 27 %
 - « acteur » « Clint Eastwood »
 - « récent » « Bertrand Blier »...
- Prendre en compte les relations exprimées dans la requête utilisateur
- Distinction sélection / projection
- Autoriser des critères de sélection dans le langage de requête
 - attribut = valeur
 - comparaison (acteur : Clint Eastwood)
 - comparaison « floue » (récent : date > 2007)

Plus long terme...

- Construction des patrons
 - manuelle ?
 - automatique ?
 - en fonction de l'ontologie
 - en fonction des requêtes fréquemment posées
- Le style des phrases descriptives laisse à désirer
- Langage de requêtes
 - des mots clés vers la LN
- Réponses pertinentes
- Explication des résultats
- Passage à l'échelle