Analyzing and automatizing BGP deduction from Triple Pattern Fragment Logs

Florent MERCIER Sitraka FIDIMIHAJAMANANA Steven FONGUE

Université de Nantes

Avril 2017

Plan

Cadre de travail

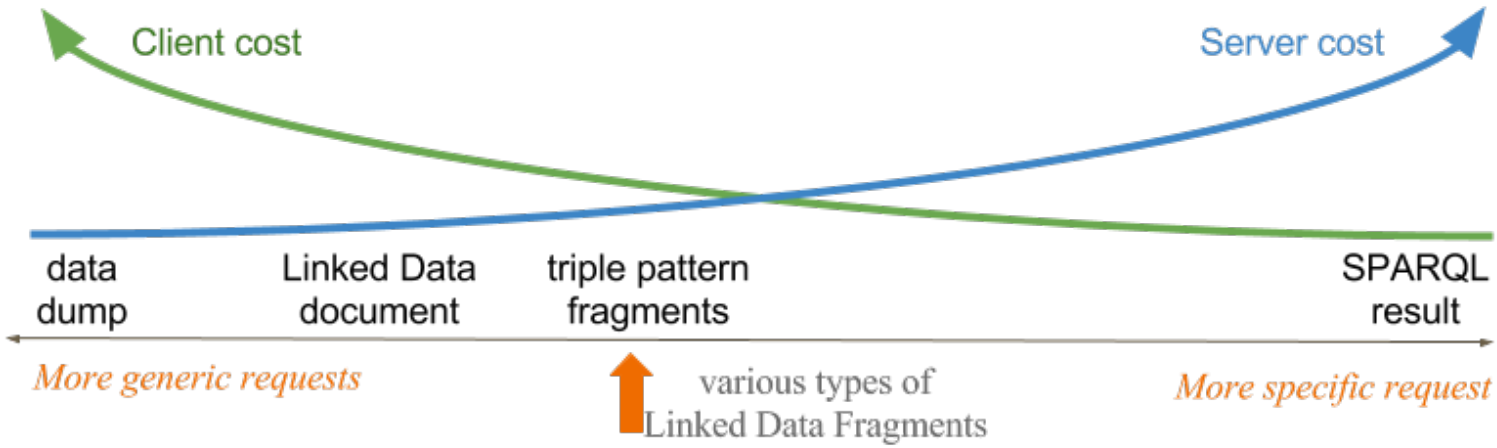
LDF (Linked Data Fragments) Ensemble de moyens pour accéder aux Linked Data

TPF (Triple Pattern Fragments) Approche permettant l'accès aux Linked Data allégant la charge du serveur

LIFT (LInked data Fragment Tracking) Outil permettant la déduction de BGP au travers d'échanges entre client/serveur TPF

BGP (Basic Graph Pattern) Ensemble de Triple Patterns

Triple Pattern Fragments (TPF)



Ruben Verborgh et al. (2016). Triple Pattern Fragments: A low-cost knowledge graph interface for the Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 37, 184-206.

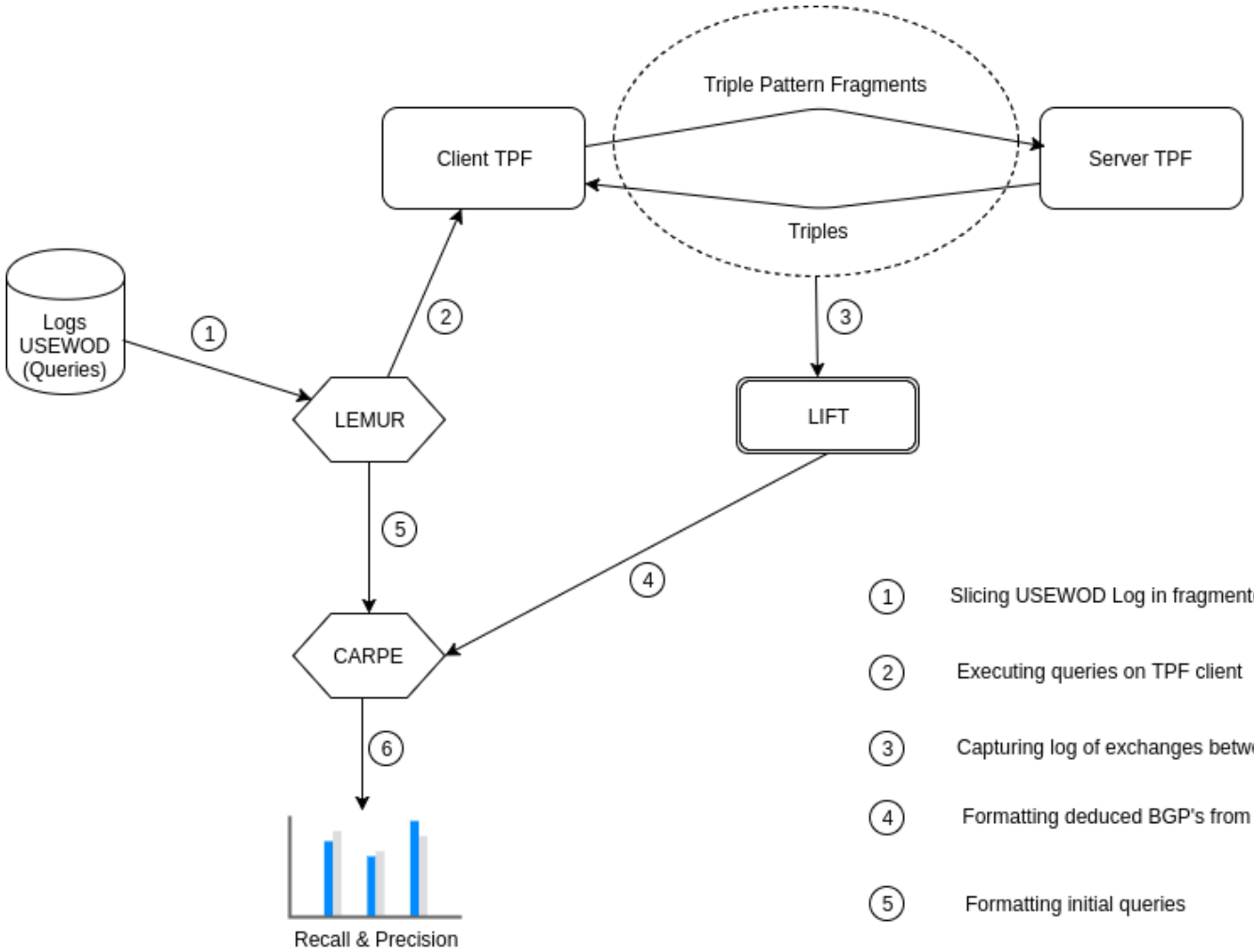
LIFT

$f(E(Q_i)) \approx BGP(Q_i)$

- f() Algorithme de LIFT
- E() Trace d'une requête
- Q_i Requête i
- BGP() BGP correspondant à une requête

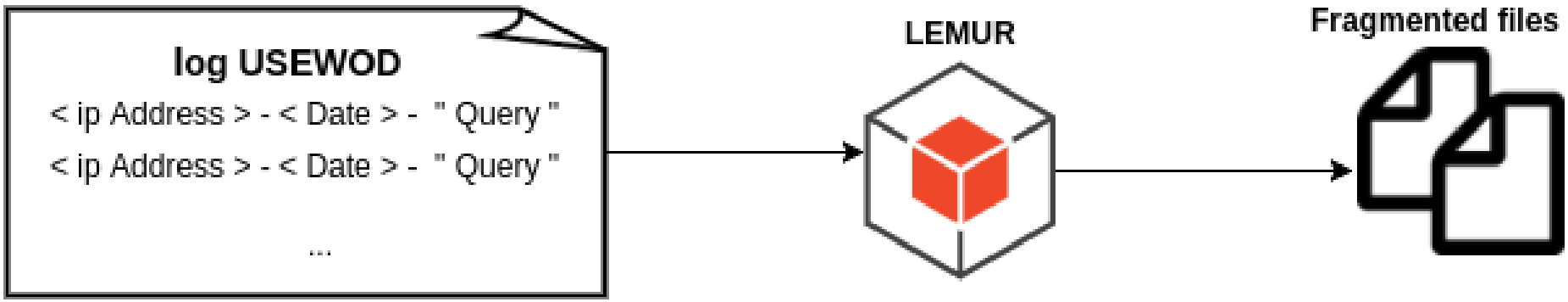
1. Détection des nested-loop
2. Recherche des inclusions
3. Extraction des BGP

Contexte



1. Slicing USEWOD Log in fragmented files
2. Executing queries on TPF client
3. Capturing log of exchanges between Client and Server
4. Formatting deduced BGP's from TPF logs
5. Formatting initial queries
6. Calculate recall and precision

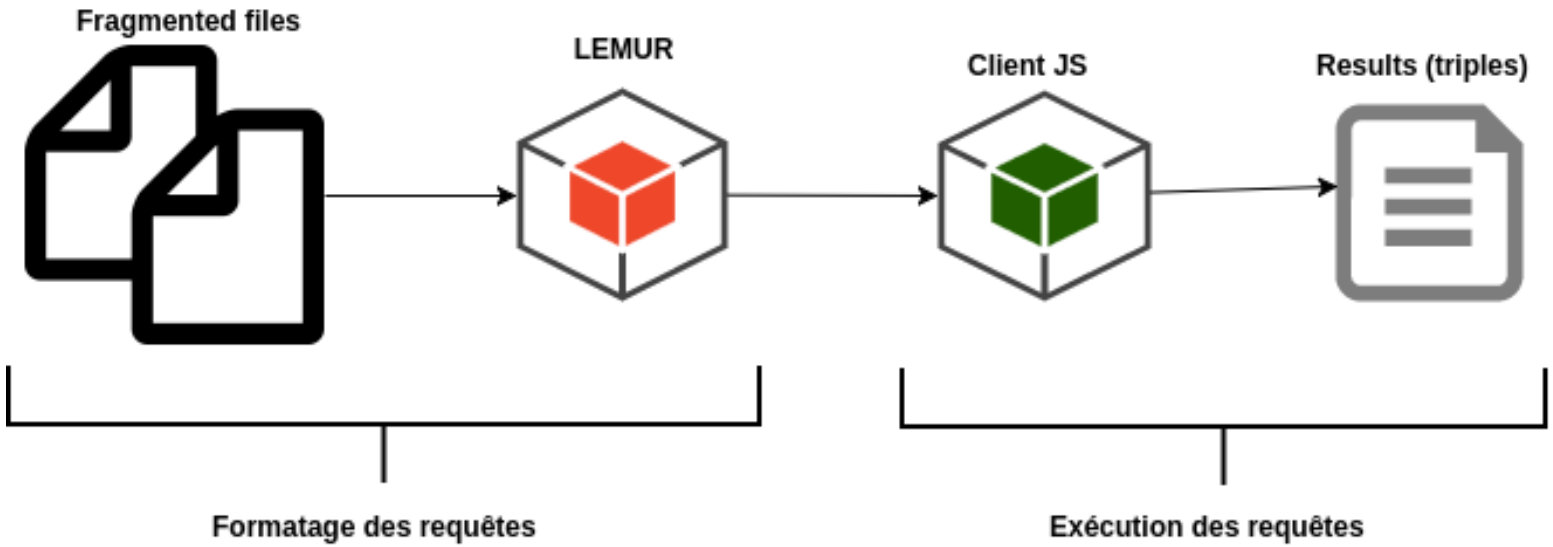
LEMUR - Découpage du log USEWOD en plusieurs fichiers



- Tri par heure
- Tri par adresse IP

Codé en JAVA

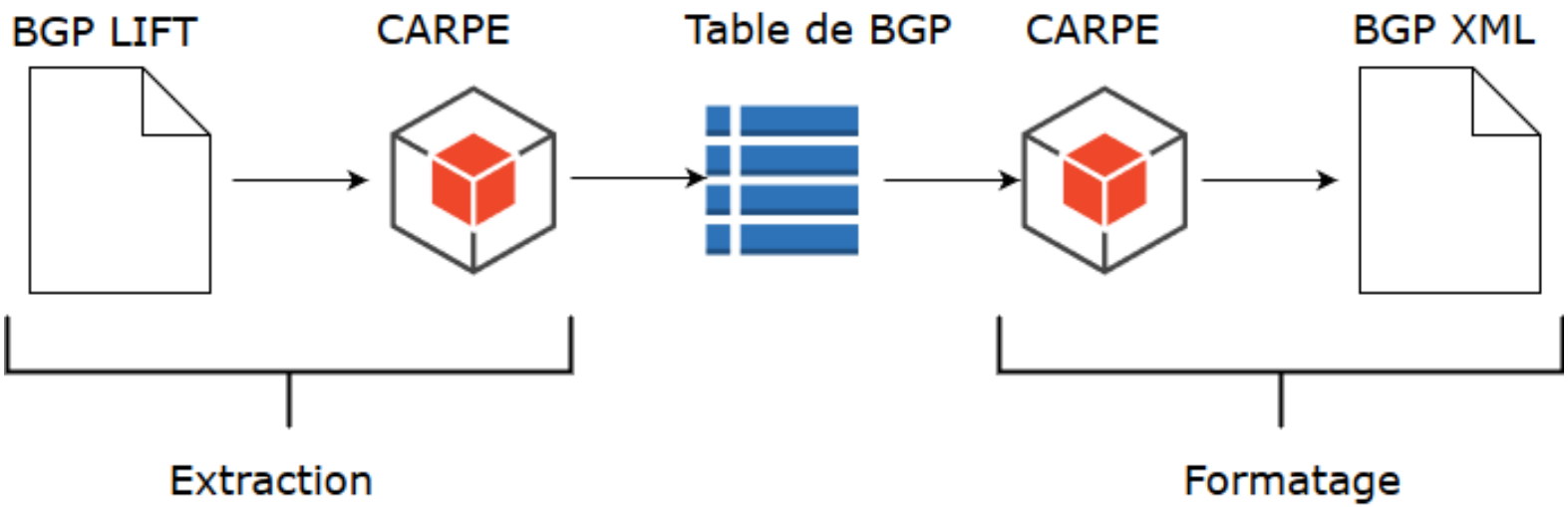
LEMUR -Formatage et exécution des requêtes à partir du client TPF



Java (Formatage des requêtes)

JavaScript (Nashorn) (Exécution des requêtes)

CARPE - Formater les BGP déduits par LIFT



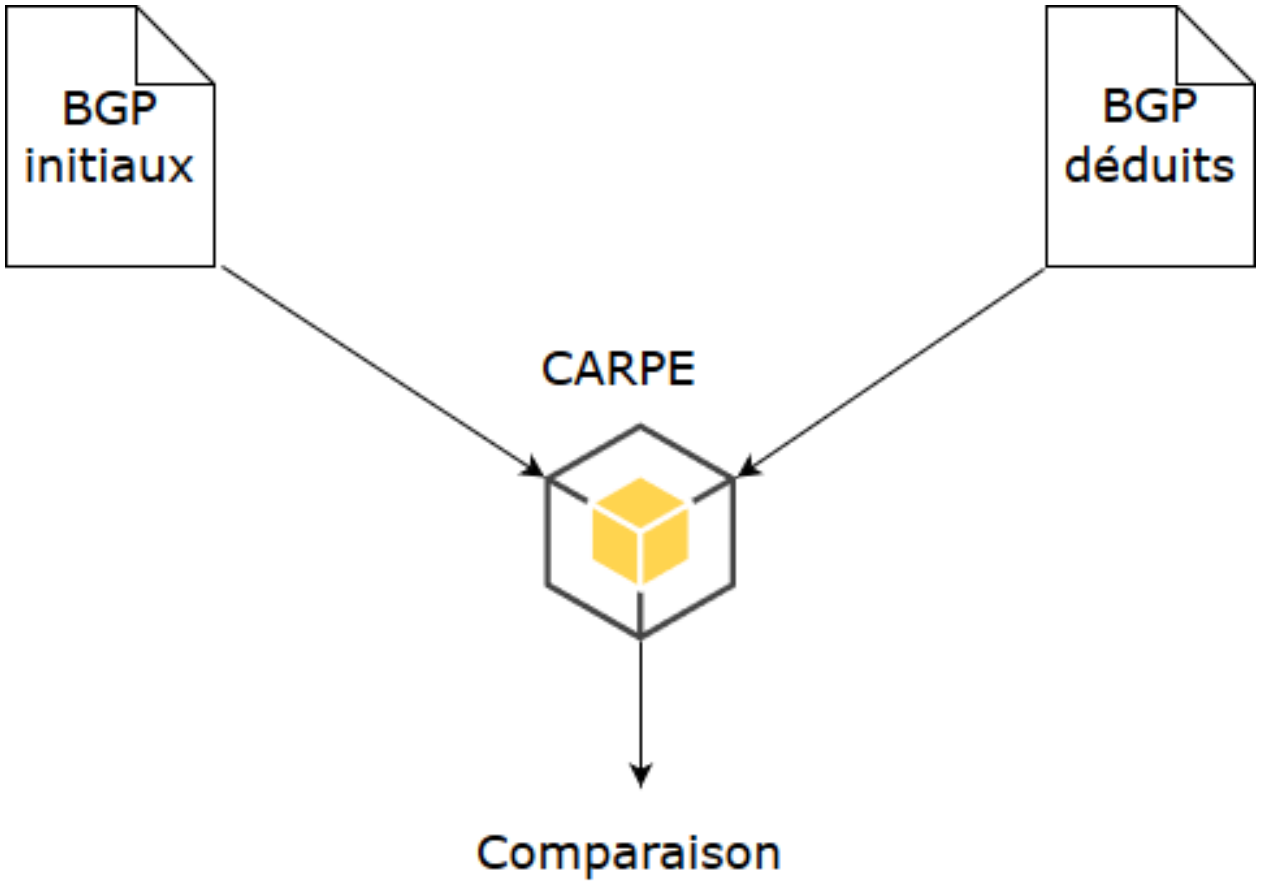
Java (Extraction des BGP)

XML (Formatage - DTD BE4DBPEDIA)

CARPE - Formater les BGP déduits par LIFT (2)

```
<?xml version="1.0" encoding="UTF-8"?>
<log>
<entry>
        <request>Select * Where {
        ?subject2
        <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
        <http://xmlns.com/foaf/0.1/Person> .
        ?subject2
        <http://xmlns.com/foaf/0.1/isPrimaryTopicOf>
        ?object3 . }</request>
        <bgp>
                <tp>?subject2
                <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
                <http://xmlns.com/foaf/0.1/Person> . </tp>
                <tp>?subject2
                <http://xmlns.com/foaf/0.1/isPrimaryTopicOf>
                ?object3 . </tp>
        </bgp>
</entry>
...
</log>
```

CARPE - Comparer ces BGP aux requêtes initiales



Java (Formatage N-Triple)

Python (Comparaison - RDFLib)

Expérimentations - LEMUR

```
8c390a2329b64fdea9b9c782107f064c - - [02/Nov/2015 03:00:00 +0100]
"GET /sparql?query=%20SELECT%20%3Fv%0A%20WHERE%20%7B%20%20%3Chttp%3A%2F%2Fdbpedia.org
%2Fresource%2FMississippi%3E%20%3Chttp%3A%2F%2Fwww.w3.org%2F2000%2F01%2Frdf-
schema%23comment%3E%20%3Fv%20.%20%7D&maxrows=10000000 HTTP/1.1" 200 5896 "-" "R" "-"
```

```
SELECT ?v
WHERE {
    <http://dbpedia.org/resource/Mississippi>
    <http://www.w3.org/2000/01/rdf-schema#comment>
    ?v .
}
```


```
[
{"?v":"\"Mississippi /ˌmɪsɨ'sɪpi/ is a state located in the Southern United States.The nam
e of the state derives from the Mississippi River, which flows along its western boundary
and comes from the Ojibwe word misi-ziibi (\"Great River\"). Jackson is the state capital
and largest city, with a population of around 175,000 people. The state overall has a popu
lation of around 3 million people. Mississippi is the 32nd most extensive and the 31st mos
t populous of the 50 United States.\"@en"}
]
```


Traitement sur le fichier USEWOD 2016 (>1 million de requêtes)

Expérimentations - LIFT

```
----------Deduced BGPs----------

    BGP [no1] +
        Deduced LDF_1: subject     http://www.w3.org/2000/01/rdf-schema#label     "Brad Pitt"@en
            received @[DBpedia 2015]
        Deduced LDF_2: subject     http://dbpedia.org/ontology/starring     INJECTEDsubj(LDF_1)
            received @[DBpedia 2015]
        Deduced LDF_6: INJECTEDsubj(LDF_2)     http://www.w3.org/2000/01/rdf-schema#label     object
            received @[DBpedia 2015]
        Deduced LDF_7: INJECTEDsubj(LDF_2)     http://dbpedia.org/ontology/director     object
            received @[DBpedia 2015]
        Deduced LDF_8: INJECTEDobj(LDF_7)     http://www.w3.org/2000/01/rdf-schema#label     object
            received @[DBpedia 2015]

----------Deduced BGPs----------
```

Expérimentations sur l'ensemble des requêtes du client web de LDF

Expérimentations - CARPE

```
<?xml version="1.0" encoding="UTF-8"?>
<log>
<entry>
        <request>Select * Where {
        ?subject2
        <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
        <http://xmlns.com/foaf/0.1/Person> .
        ?subject2
        <http://xmlns.com/foaf/0.1/isPrimaryTopicOf>
        ?object3 . }</request>
        <bgp>
                <tp>?subject2
                <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
                <http://xmlns.com/foaf/0.1/Person> . </tp>
                <tp>?subject2
                <http://xmlns.com/foaf/0.1/isPrimaryTopicOf>
                ?object3 . </tp>
        </bgp>
</entry>
...
</log>
```

Essai concluant sur un ensemble de 3000 fichiers de déduction

Expérimentations - Comparaison de graphes

```
graphs are isomorphic ?
True
in both :
_:cb9c7fe2f0677a92478a080ef9247b0a7b39b4c179521c2cec69377c131ca34bc5 <http://www.w3.org/2000/0
1/rdf-schema#label> _:cb0 .
_:cbbfb403114937179c284895c1d9138551fd34bf4f742947b6bc66e5216522dac5 <http://dbpedia.org/ontol
ogy/director> _:cb9c7fe2f0677a92478a080ef9247b0a7b39b4c179521c2cec69377c131ca34bc5 .
_:cbbfb403114937179c284895c1d9138551fd34bf4f742947b6bc66e5216522dac5 <http://dbpedia.org/ontol
ogy/starring> _:cbe024b43504549359da613f5927a624843065805989 88d41757a648eebf4d0395 .
_:cbbfb403114937179c284895c1d9138551fd34bf4f742947b6bc66e5216522dac5 <http://www.w3.org/2000/0
1/rdf-schema#label> _:cbdd9e86c14337f1edc712125a155dcabc1d6cba8f892cef4aa3d94e47ccfcb810 .
_:cbe024b43504549359da613f5927a624843065805989 88d41757a648eebf4d0395 <http://www.w3.org/2000/0
1/rdf-schema#label> "Brad Pitt"@en .
in first :
in second :
```

Tests sur des requêtes isolées

Conclusion

Non fonctionnel

- Analyse et récupération des échanges entre le client et le serveur TPF
- Calcul de la précision et du rappel

Fonctionnel mais non utilisé

- Fragmentation du log USEWOD
- Exécution automatique des requêtes originelles
- Comparaison des graphes

Fonctionnel et utilisé

- Extraction des BGP en sortie de LIFT
- Formatage XML des requêtes déduites