SWEEP: a Streaming Web Service to Deduce Basic Graph Patterns from Triple Pattern Fragments

Emmanuel Desmontils, Patricia Serrano-Alvarado and Pascal Molli
LS2N - University of Nantes – France
{firstname.lastname}@univ-nantes.fr

Motivation

It is fundamental to understand user queries executed over servers **to detect recurrent patterns in queries** for prefetching, auditing, usage control, etc.

Currently, this is not possible in the context of TPF (Triple Pattern Fragments) because servers only receive single triple patterns.

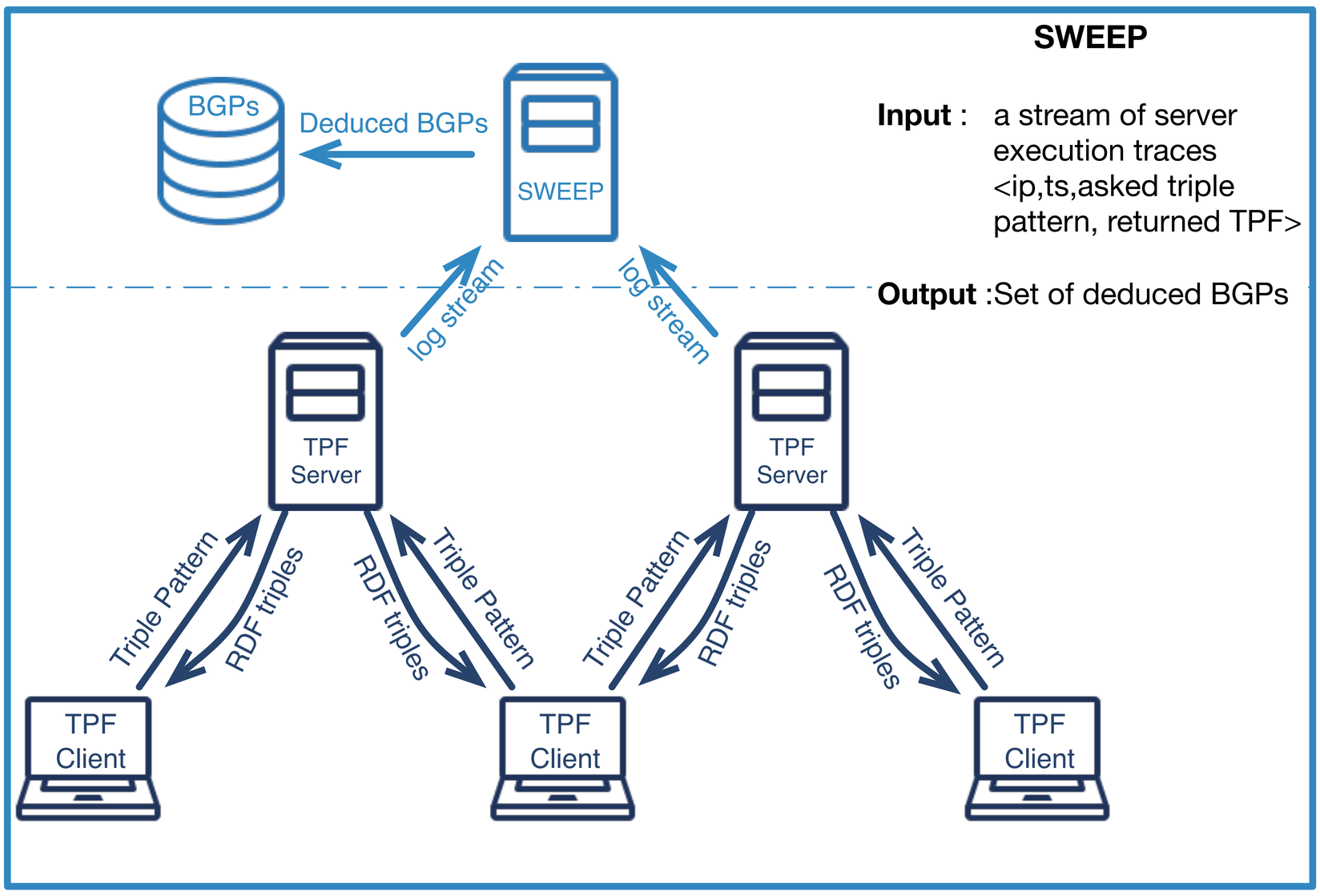
Objectives

To deduce BGPs (Basic Graph Patterns) of user's queries from TPF server logs.

Main challenge: to distinguish concurrent similar SPARQL queries.

Approach

From a stream of execution traces organized by IP-address and a sliding window, SWEEP deduces triple pattern joins implemented through nested-loops.



SWEEP

Input : a stream of server execution traces <ip,ts,asked triple pattern, returned TPF>

Output : Set of deduced BGPs

SWEEP supports:

- 🙂 TPF pagination
- 🙂 Multiple TPF server logs
- 🙂 Federated queries
- 😐 Concurrency

Evaluation

SWEEP produces a subset of the partition of the BGP of the original SPARQL query, i.e., the query execution plan. SWEEP was evaluated with the set of queries of the public TPF client.
SWEEP obtained 100% of precision and 87% of recall when the best deduced BGP is compared to the original BGP.

SWEEP TPF Client

QUERY EDITOR

Base : dbpedia3.8

Enter the query :

```
PREFIX dbpedia-owl: <http://dbpedia.org/ontology/>
PREFIX dbpprop: <http://dbpedia.org/property/>
SELECT DISTINCT ?entity WHERE {
?entity a dbpedia-owl:Airport;
        dbpprop:cityServed dbpedia:Italy.
}
```

Reference query. The BGP is handmade and validated.

	?entity
1	http://dbpedia.org/resource/Albenga_Airport
2	http://dbpedia.org/resource/Ancona_Airport
3	http://dbpedia.org/resource/Belluno_Airport
4	http://dbpedia.org/resource/Casale_Airport
5	http://dbpedia.org/resource/Centocelle_Airport
6	http://dbpedia.org/resource/Foggia_%22Gino_Lisa%22_Airport
7	http://dbpedia.org/resource/Forl%C3%AC_Airport
8	http://dbpedia.org/resource/Galileo_Galilei_Airport
9	http://dbpedia.org/resource/Grosseto_Airport
10	http://dbpedia.org/resource/L'Aquila-Preturo_Airport
11	http://dbpedia.org/resource/Naples_Airport
12	http://dbpedia.org/resource/Palermo_Airport
13	http://dbpedia.org/resource/Parma_Airport
14	http://dbpedia.org/resource/Perugia_San_Francesco_d'Assisi_%E2%80%93_Umbria_International_Airport
15	http://dbpedia.org/resource/Siena-Ampugnano_Airport
16	http://dbpedia.org/resource/Taranto-Grottaglie_Airport
17	http://dbpedia.org/resource/Turin_Caselle_Airport
18	http://dbpedia.org/resource/Venice_Marco_Polo_Airport

(0.479636 sec.)

SWEEP dashboard

BGP under construction: 0 | Queries to find: 0 | Avg Precision: 100% | Avg Recall: 100%

INFORMATION

Gap (hh:mm:ss)	Evaluated Queries	BGP	TPQ
0:00:15	2 / 2	7	144

DEDUCED BGPS

(20 more recents)

	ip	time	bgp	Original query	Precision	Recall
7	193.52.103.60	2017-10-05 11:57:39.785151	?js0 <http://dbpedia.org/property/cityServed> <http://dbpedia.org/resource/Italy> . ?js0 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology/Airport> .	qsim-WS-193.52.103.62-2_1 PREFIX dbpedia-owl: <http://dbpedia.org/ontology/> PREFIX dbpprop: <http://dbpedia.org/property/> SELECT DISTINCT ?entity WHERE { ?entity a dbpedia-owl:Airport; dbpprop:cityServed dbpedia:Italy. }	1.000	1.000
6	193.52.103.60	2017-10-05 11:57:40.133688	?se26789_134 <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://dbpedia.org/ontology /Airport> .	No query assigned		
5	193.52.103.60	2017-10-05 11:56:09.414510	?jo0 <http://www.w3.org/2000/01/rdf-schema#label> "Brad Pitt"@en . ?js1 <http://dbpedia.org/ontology/starring> ?jo0 . ?js1 <http://www.w3.org/2000/01/rdf-schema#label> ?oe26853_142 . ?js1 <http://dbpedia.org/ontology/director> ?js3 . ?js3 <http://www.w3.org/2000/01/rdf-schema#label> ?oe26853_143 .	qsim-WS-193.52.103.62-1_1 PREFIX dbpedia-owl: <http://dbpedia.org/ontology/> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> SELECT ?movie ?title ?name WHERE { ?movie dbpedia-owl:starring [rdfs:label "Brad Pitt"@en]; rdfs:label ?title; dbpedia-owl:director [rdfs:label ?name] }	1.000	1.000
4	193.52.103.60	2017-10-05 11:56:09.829976	?se26795_751 <http://www.w3.org/2000/01/rdf-schema#label> ?oe26795_751 .	No query assigned		
3	193.52.103.60	2017-10-05 11:56:09.803813	?se26853_140 <http://www.w3.org/2000/01/rdf-schema#label> ?oe26853_140 .	No query assigned		