

ISWC2016の報告

Hitachi, Ltd.

Research & Development Group

Yu Asano and Makoto Iwayama

ISWC2016 自然言語処理・QA・統計

【傾向】

- Knowledge Graphを使った論文が多かった。
- 自然言語処理では、Resource Trackの論文が多かった。セマンティックWeb技術を用いた言語処理アプリケーションの実用に向けた環境が整いつつある。
- 統計分野の発表は、これまでワークショップ中心だったが、本会議でも見られるようになってきた。

【紹介する発表】

1. QuerioDALI: Question Answering over Dynamic And LIinked knowledge graphs
Vanessa Lopez, Pierpaolo Tommasi, Spyros Kotoulas and Jiewen Wu (IBM)
2. CubeQA- Question Answering on RDF Data Cubes
Konrad Hoffner, Jens Lehmann and Ricardo Usbeck (University of Leipzig)
3. LODStats: The Data Web Census Dataset
Ivan Ermilov, Jens Lehmann, Michael Martin and Soren Auer
(University of Leipzig)

1. QuerioDALI: Question Answering over Dynamic And Linked knowledge graphs

1-1 背景

- LODやKGの普及により、構造化データ(知識)を用いたQAが増加
cf.) 従来のQAの多くはテキスト検索ベース
- テストコレクションも増加
 - QALD: DBpediaを対象にしたQA
 - Free917: Freebaseを対象にしたQA

1-2 課題

- 知識が増えてきた反面、それらを効率良く扱うことが必要
→ Watsonを用いたpracticalなアプローチ
- 複数の知識源(KG)を統合することが必要。
ex.) “What are the side-effects for all the medications of this patient?”
(ある患者が投与している薬(病院内のデータ)とその薬の副作用(オープンデータ)を組み合わせないと答えられない)
→ テンプレート(グラフパターン)の統合

1. QuerioDALI: Question Answering over Dynamic And Linked knowledge graphs

1-3 提案

1. 構文解析によりQから述語項構造(SVO)を抽出
Watson deep parseおよびNE(固有名抽出)を使用
2. 述語項構造の各要素をURIに変換
ラベルとの表記類似度およびWordNet等による展開
3. グラフパタンの生成
18種類のQの型に応じたテンプレートを使用
ex.) “highest mountain in Australia”に対して

```
{?s a db:Mountain. ?s ?px db:Australia. ?s db:elevation ?o}  
ORDERBY DESC(?0) LIMIT 1 OFFSET 0
```


を生成
4. グラフパタンの統合
通常のjoinに加えsyntectic(ラベルの類似性)/semantic(同値リンク)
な関係も考慮
5. 検索およびAのランキング

1. QuerioDALI: Question Answering over Dynamic And Linked knowledge graphs

1-4 効果

- オープンドメイン
 - QALD-5(DBPediaに対するQA)とFree917(Freebaseに対するQA)で評価しトップレベルのF値
- 医療ドメイン
 - QALD-4で評価
 - 複数のグラフを統合しないと答えられないQがある
 - 他のシステムとの比較は無い

2. CubeQA- Question Answering on RDF Data Cubes

2-1 背景

- 世界的にオープンデータ公開が推進
- 統計データ（多次元データ）はヘルスケア、政策、金融などの分野で活用可能
- Semantic Question Answering (SQA)技術により、自然言語クエリによる問い合わせが可能

2-2 課題

- SQAは統計データ（RDF Data Cubeで表現）には適用できない

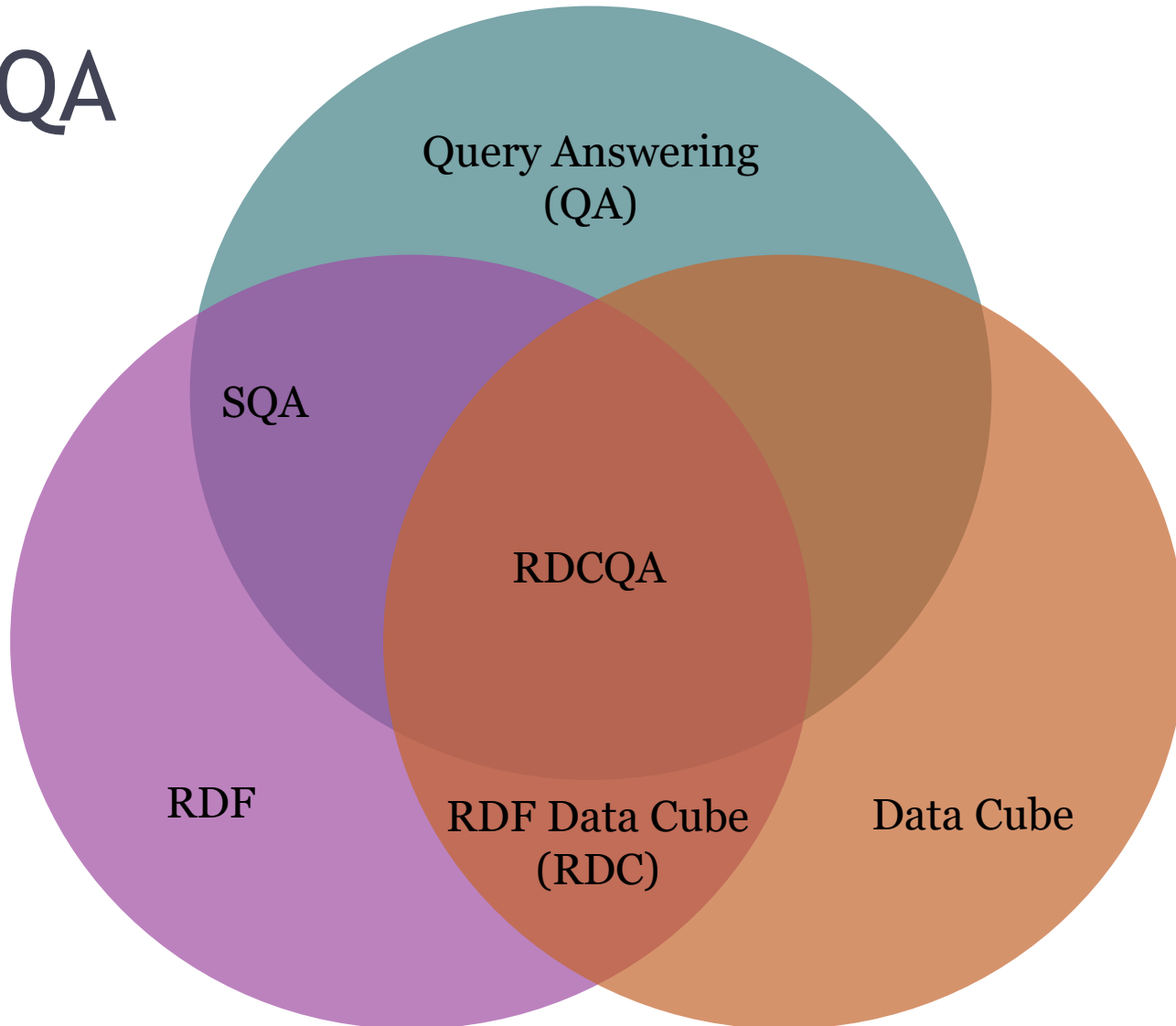
2-3 提案

- RDCQAをRDF Data Cubeに対するSQAとして定義し、RDCQAシステムのためのベンチマーク（QALD6T3-test）を作成
- **CubeQA algorithm**を提案

2-4 効果

- CubeQA algorithmを提案ベンチマークで評価した結果、F値が0.43となり、提案システムの実用可能性を示した。

RDCQA



CubeQA algorithmの紹介

- 自然言語の質問をSPARQLに変換するためのアルゴリズム

ex.)

自然言語の質問

How much did the
Philippines receive in
the year of 2007?

変換

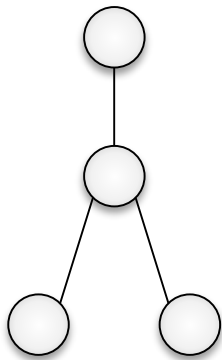
SPARQL

```
SELECT sum(?amount)
FROM :finaland-aid
{
  ?o a qa:Observation.
  ?o :amount ?amount.
  ?o :recipient-country :ph.
  ?o :refYear ?y.
  filter (year(?y)=2007)
}
```


CubeQA algorithmの紹介

- 自然言語の質問をSPARQLに変換するためのアルゴリズム

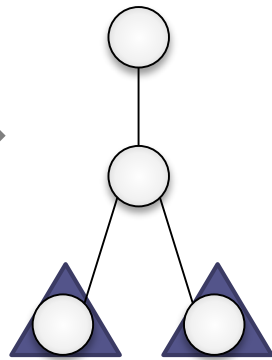
1. Preprocessing



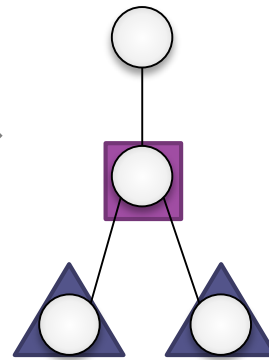
自然言語の質問

How much did the Philippines receive in the year of 2007?

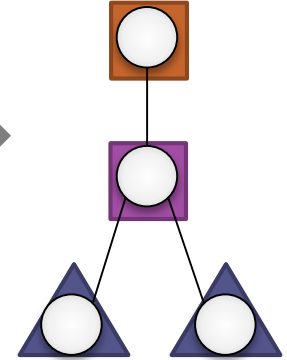
2. Matching



3. Combination



4. Execution



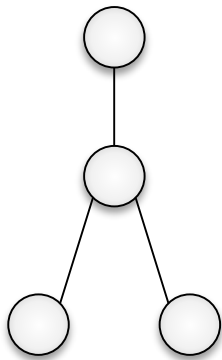
SPARQL

```
SELECT ...  
FROM ...  
{ ... }
```

CubeQA algorithmの紹介

- 自然言語の質問をSPARQLに変換するためのアルゴリズム

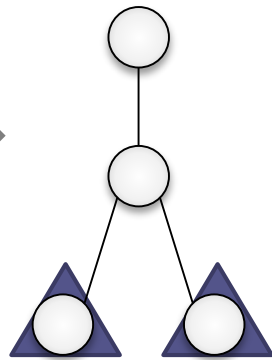
1. Preprocessing



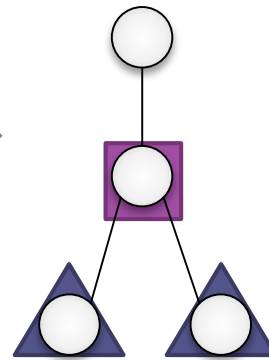
自然言語の質問

How much did the Philippines receive in the year of 2007?

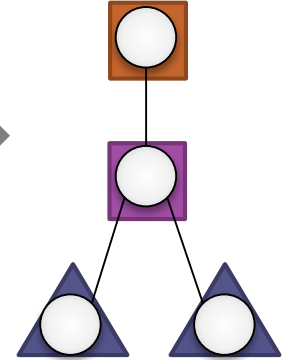
2. Matching



3. Combination



4. Execution

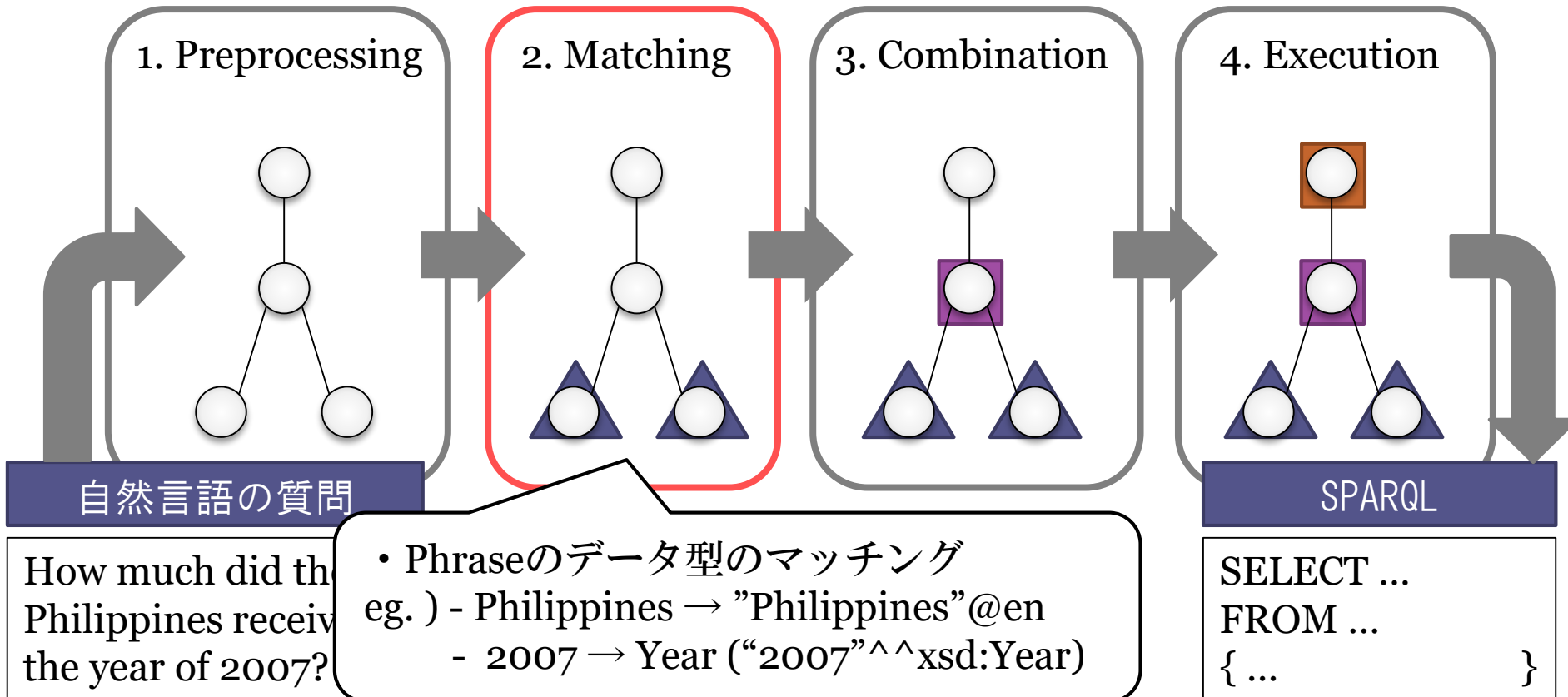


SPARQL

```
SELECT ...  
FROM ...  
{ ... }
```

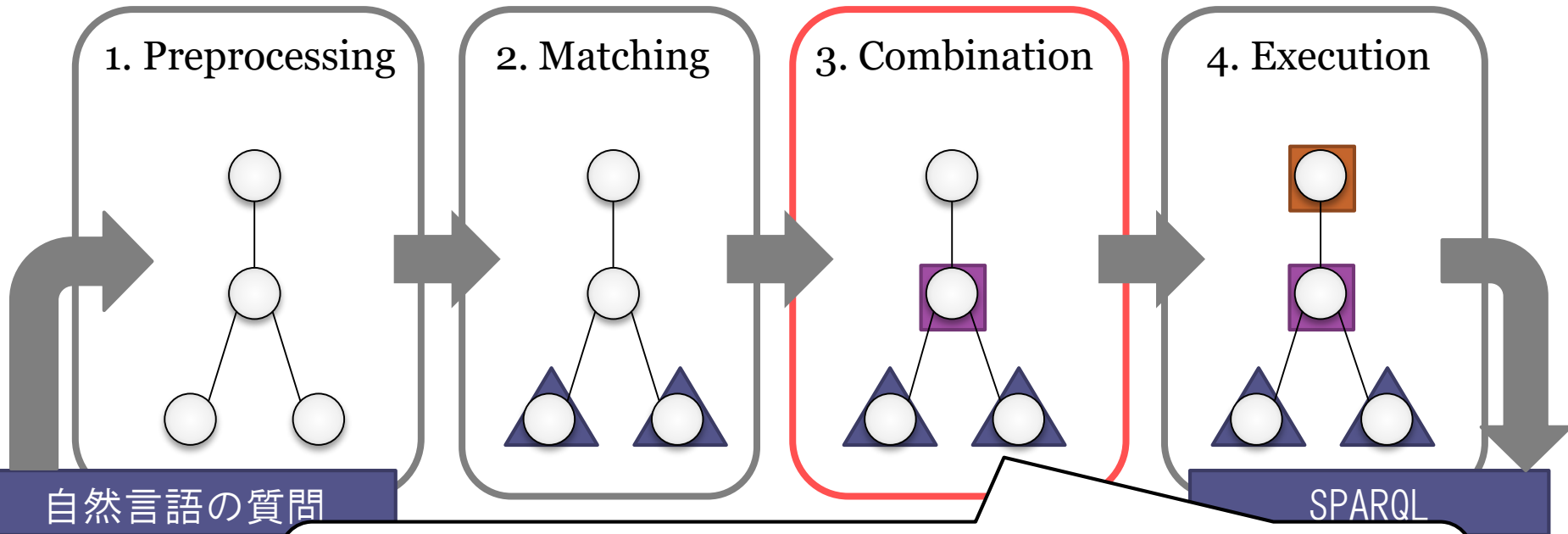
CubeQA algorithmの紹介

- 自然言語の質問をSPARQLに変換するためのアルゴリズム



CubeQA algorithmの紹介

- 自然言語の質問をSPARQLに変換するためのアルゴリズム



How much did t
Philippines rece
the year of 2007

- Constraintの生成

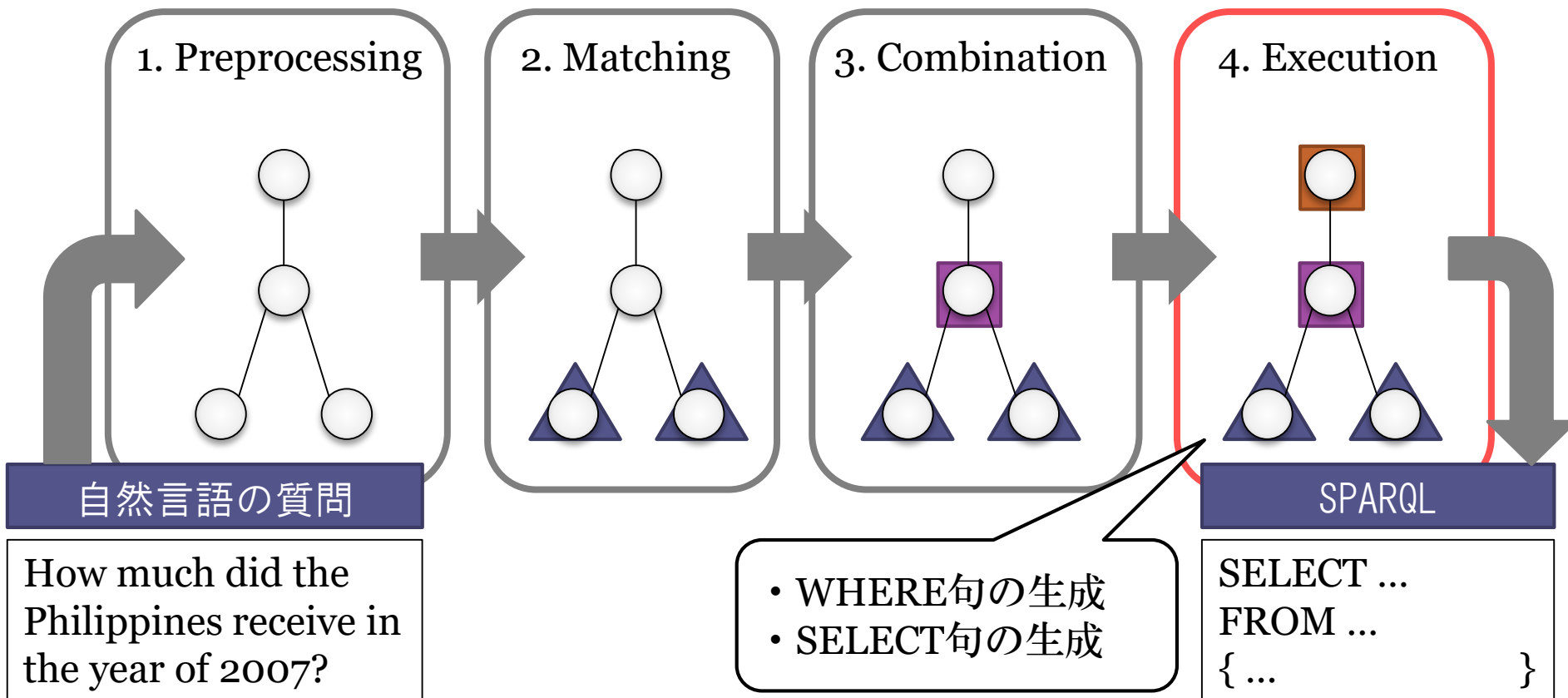
eg.) - Philippines \rightarrow `{?o :recipient-country :ph.}`

- 2007 \rightarrow `{?o :refYear ?y, filter(year(?y)=2007)}`

}

CubeQA algorithmの紹介

- 自然言語の質問をSPARQLに変換するためのアルゴリズム



3. LODStats: The Data Web Census Dataset

3.1 背景

- ここ数年で、Data Webのサイズが増加
- 多くの構造化された統計データ（RDF）も公開
ex.) data.gov
publicdata.eu
datahub.io

3.2 課題

- データの情報（質、プライバシー、カバー率等）が管理されていない

3.3 提案

- データの情報を管理するためのLODStatsを提案
- データの情報を記述するための語彙
(LDSO: LODStats DataSet vOcabulary) を提案

3.4 効果

- LODStatsのSPARQL endpointからデータの情報を用いて一括データ検索が可能