

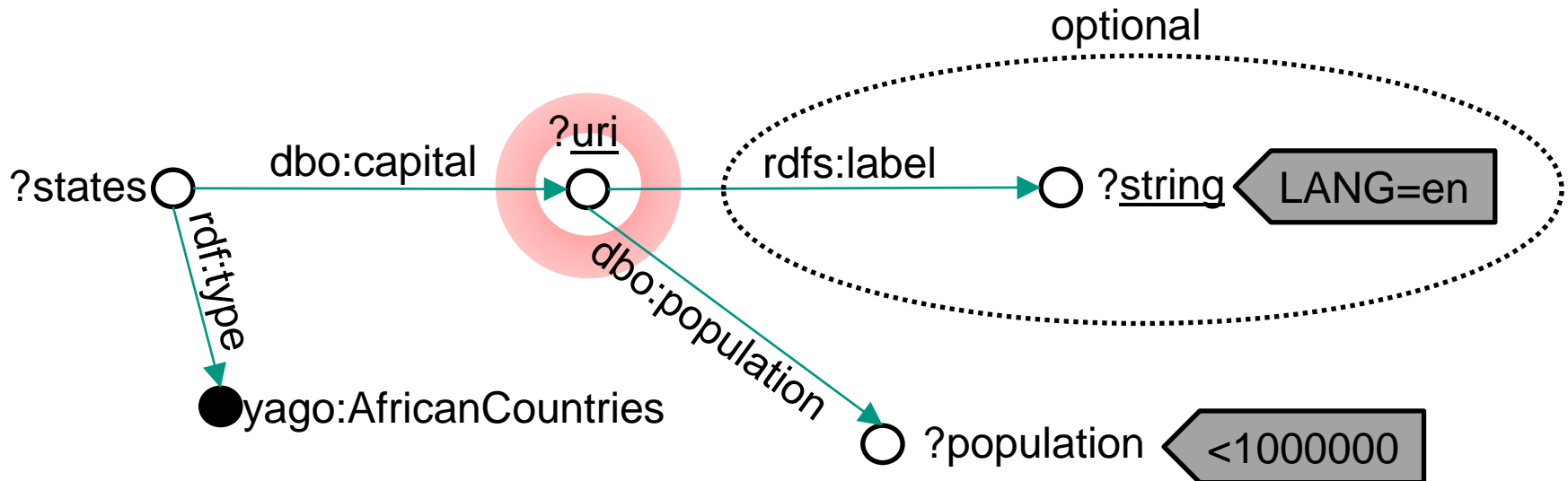
SPARTIQUULATION

Verbalizing SPARQL queries

Basil Eil, Denny Vrandečić, Elena Simperl

International Workshop on Interacting with Linked Data, Extended Semantic Web Conference 2012
28 May 2012

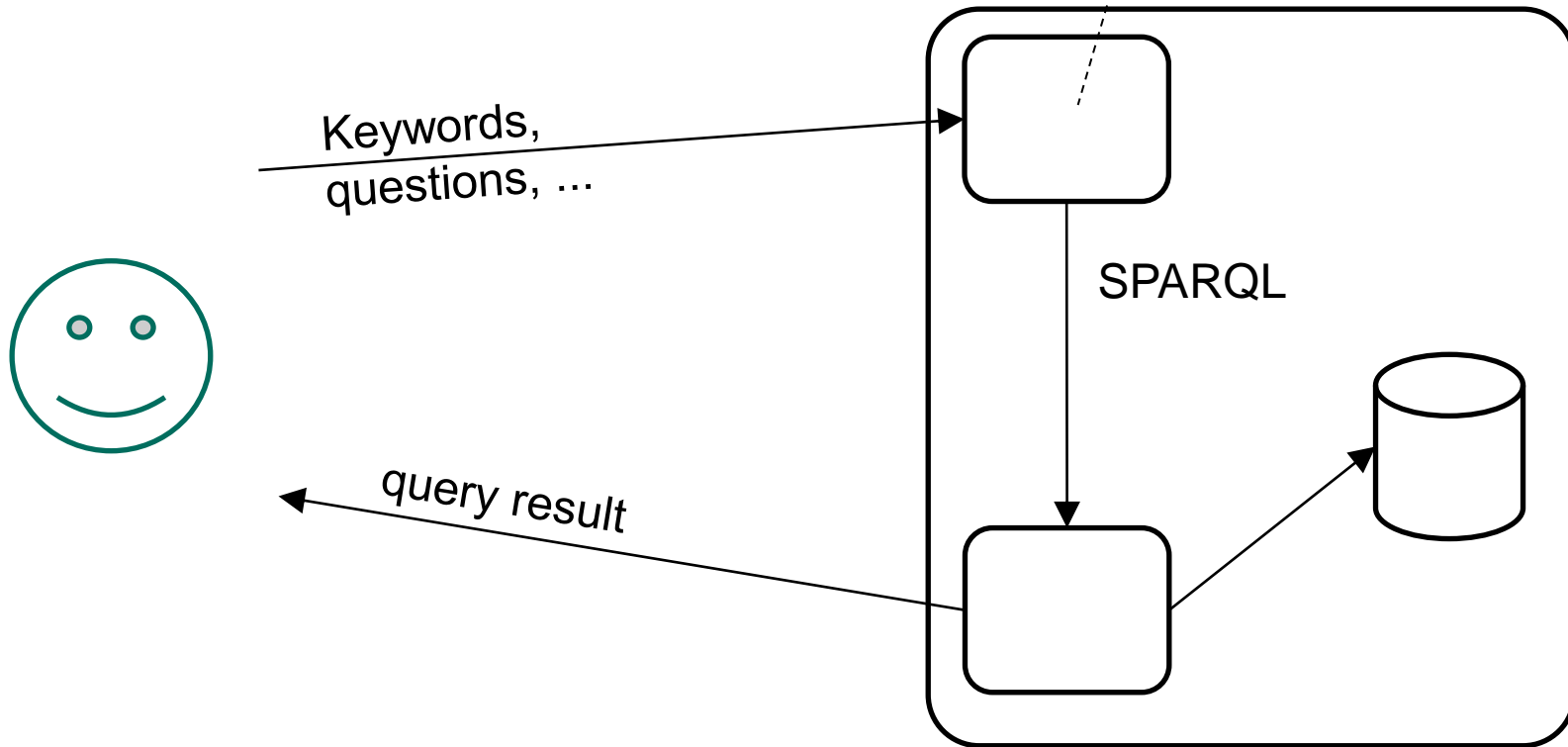
Institute of Applied Informatics and Formal Description Methods (AIFB)



MOTIVATION

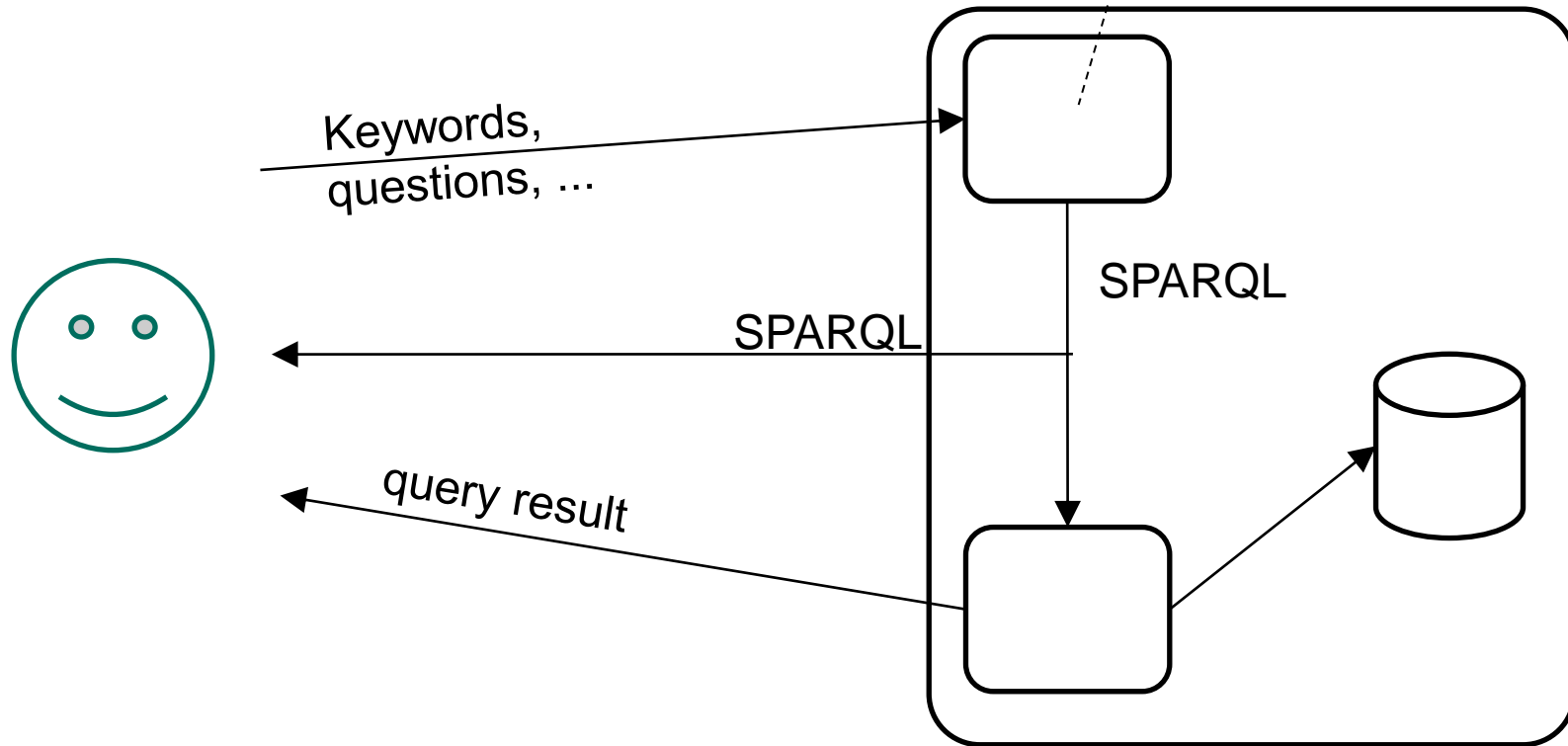
Motivation

[QALD 2011]
[Haase et al., 2009]
[Shekarpour et al., 2011]



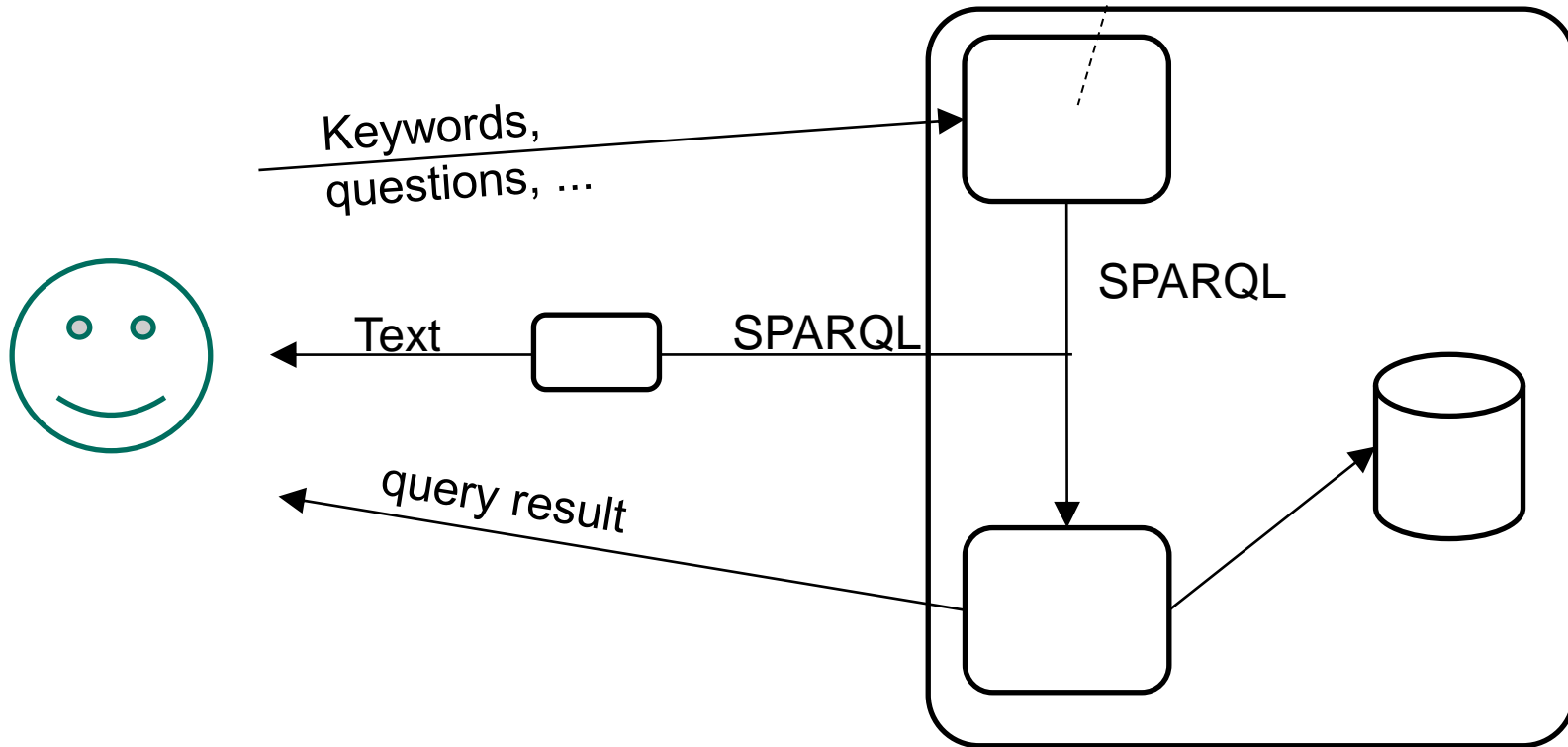
Motivation

[QALD 2011]
[Haase et al., 2009]
[Shekarpour et al., 2011]



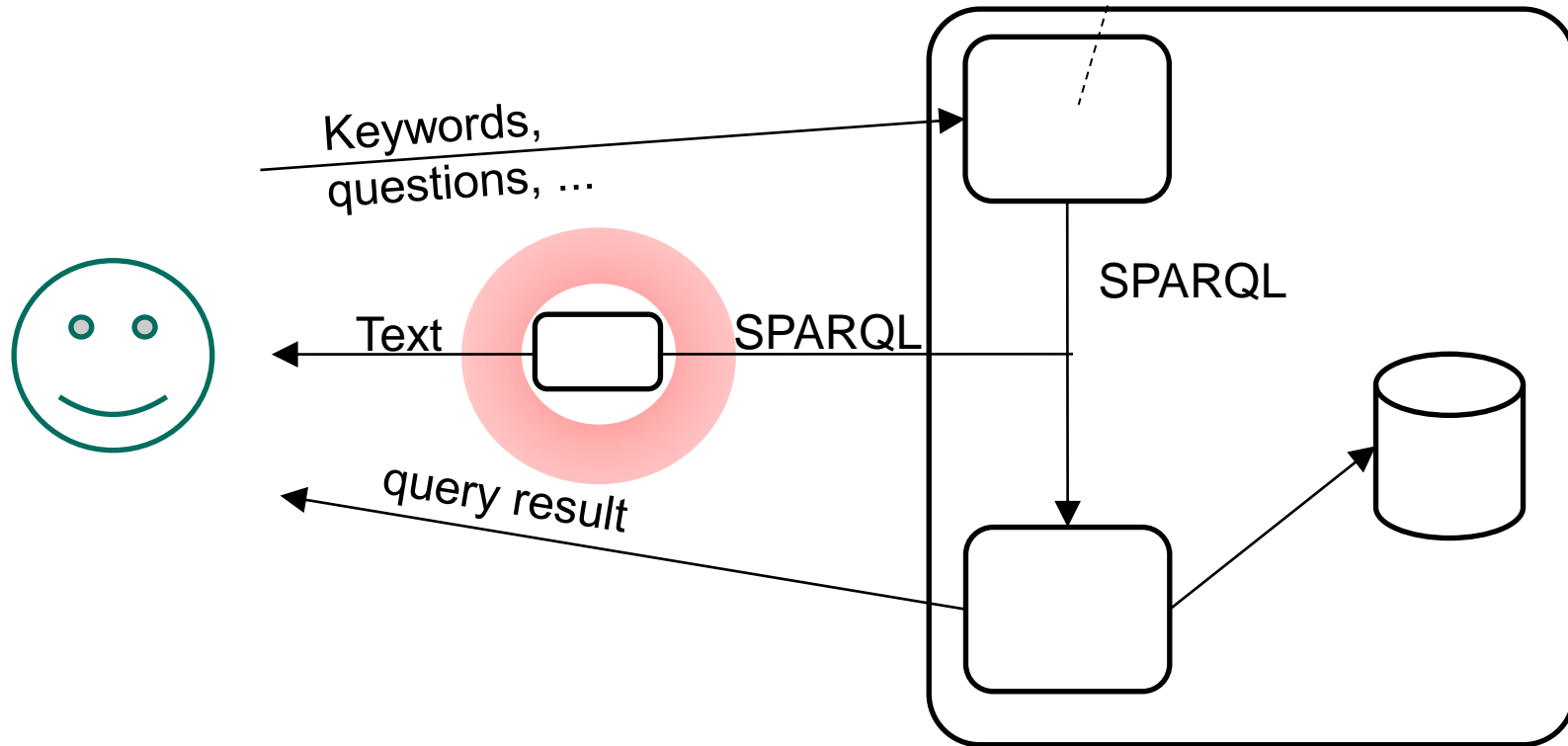
Motivation

[QALD 2011]
[Haase et al., 2009]
[Shekarpour et al., 2011]



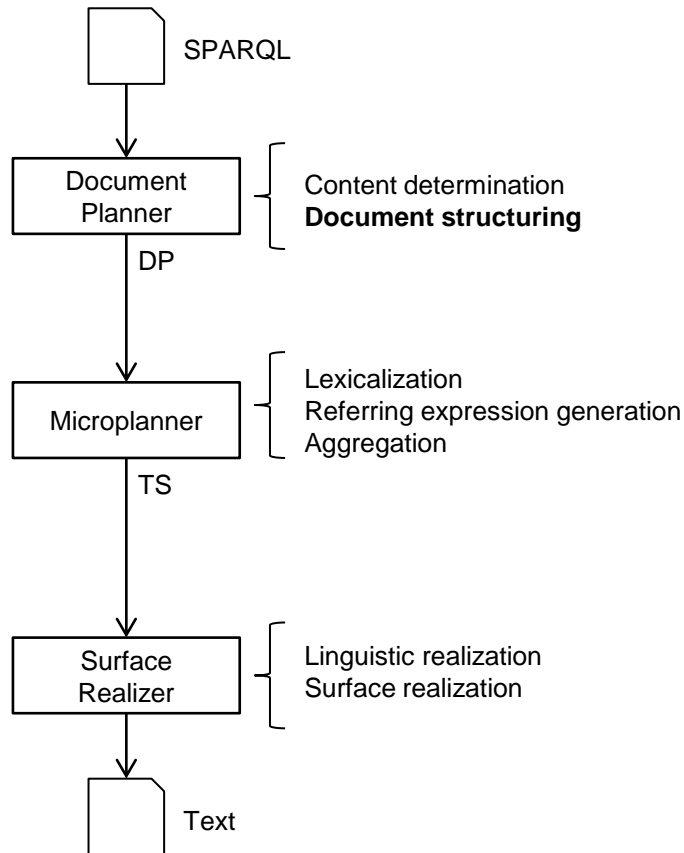
Motivation

[QALD 2011]
[Haase et al., 2009]
[Shekarpour et al., 2011]



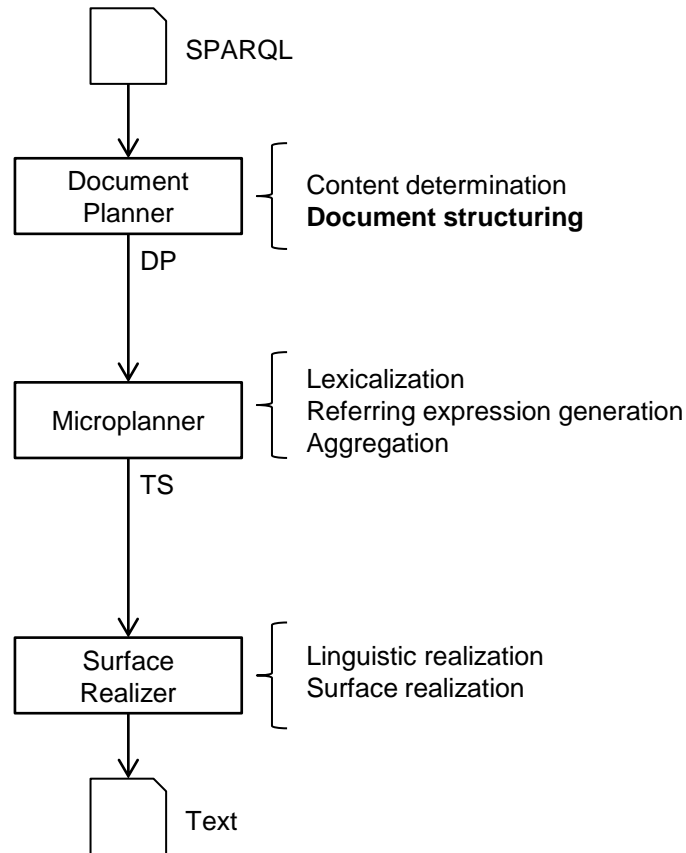
APPROACH

The pipeline architecture



[Reiter and Dale, 2000]

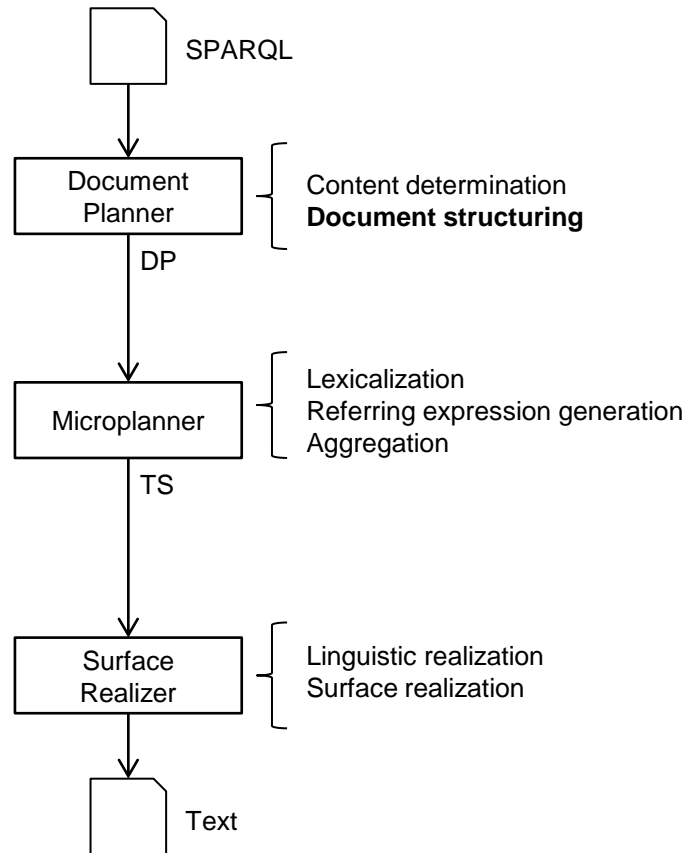
The pipeline architecture



1. Select the information to communicate
2. Constructing messages and deciding for their ordering and structure

[Reiter and Dale, 2000]

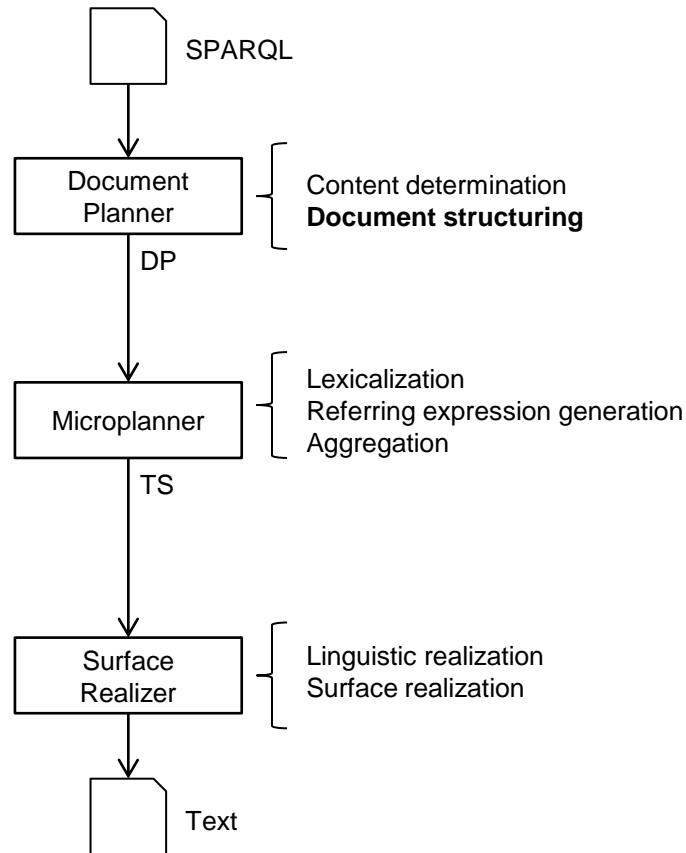
The pipeline architecture



1. Select the information to communicate
2. Constructing messages and deciding for their ordering and structure
3. Decide which words to use in order to express the content
4. Decide how to refer to an entity
5. Map to linguistic structures

[Reiter and Dale, 2000]

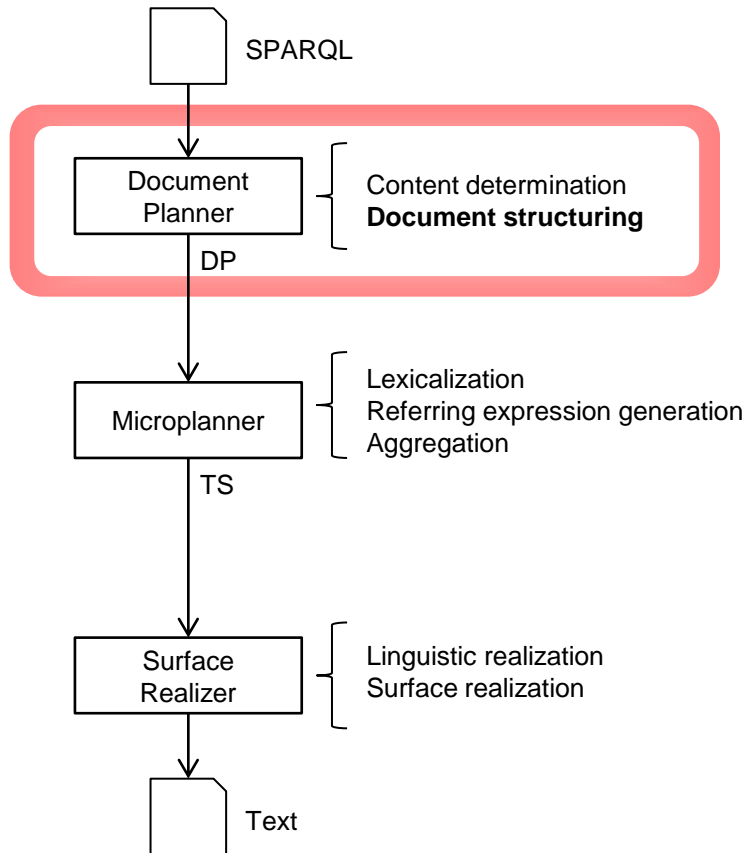
The pipeline architecture



1. Select the information to communicate
2. Constructing messages and deciding for their ordering and structure
3. Decide which words to use in order to express the content
4. Decide how to refer to an entity
5. Map to linguistic structures
6. Create natural language
7. Add structure to text such as HTML elements

[Reiter and Dale, 2000]

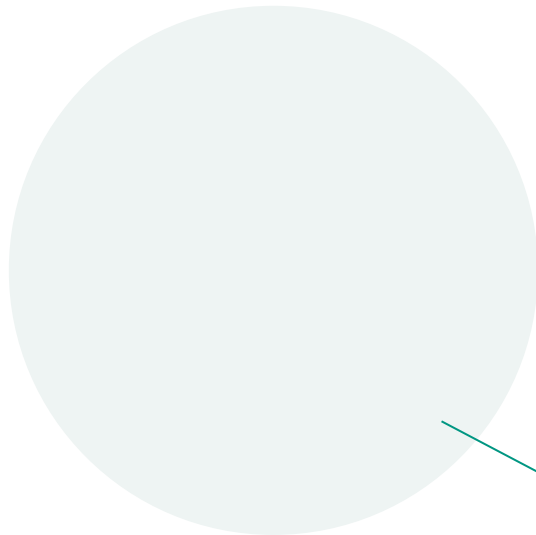
The pipeline architecture



1. Select the information to communicate
2. Constructing messages and deciding for their ordering and structure
3. Decide which words to use in order to express the content
4. Decide how to refer to an entity
5. Map to linguistic structures
6. Create natural language
7. Add structure to text such as HTML elements

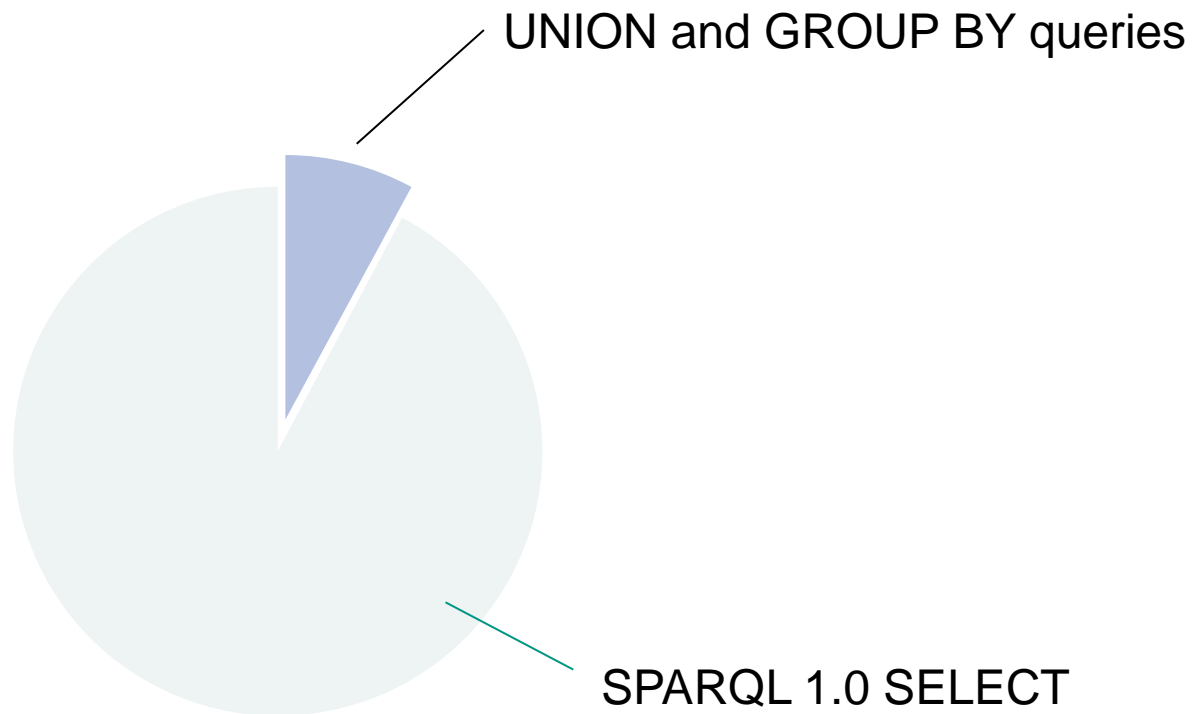
[Reiter and Dale, 2000]

Restrictions

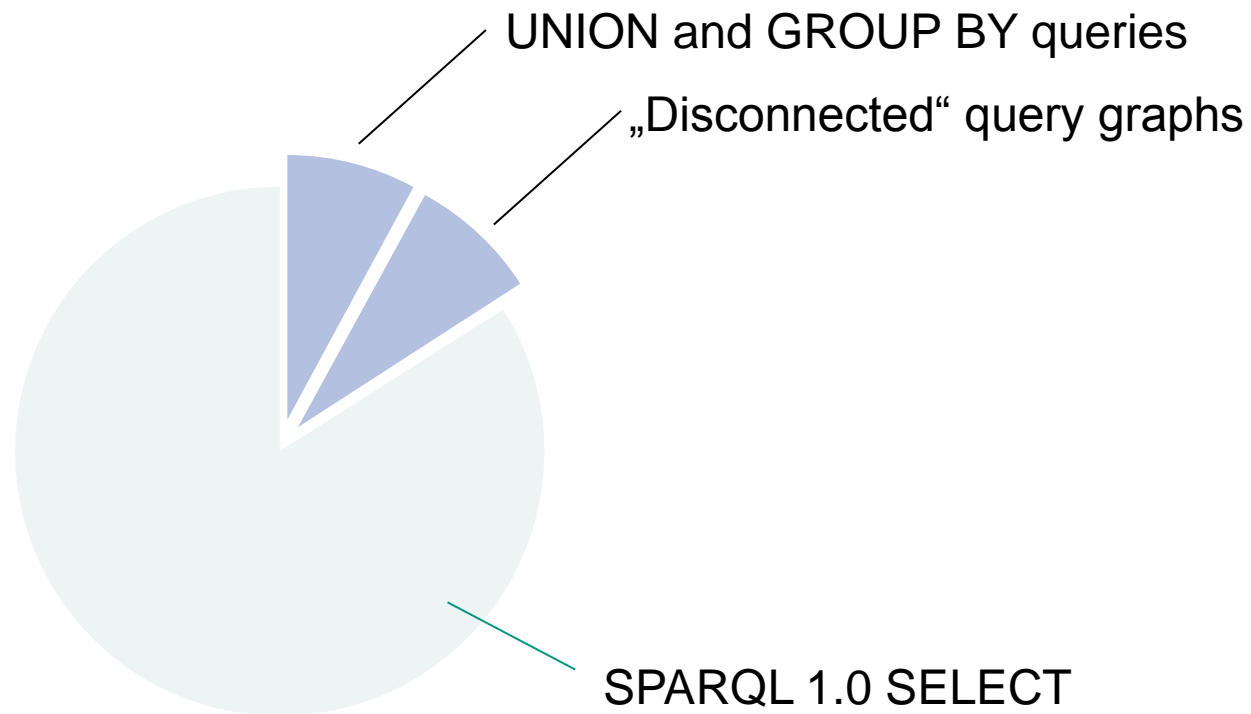


SPARQL 1.0 SELECT

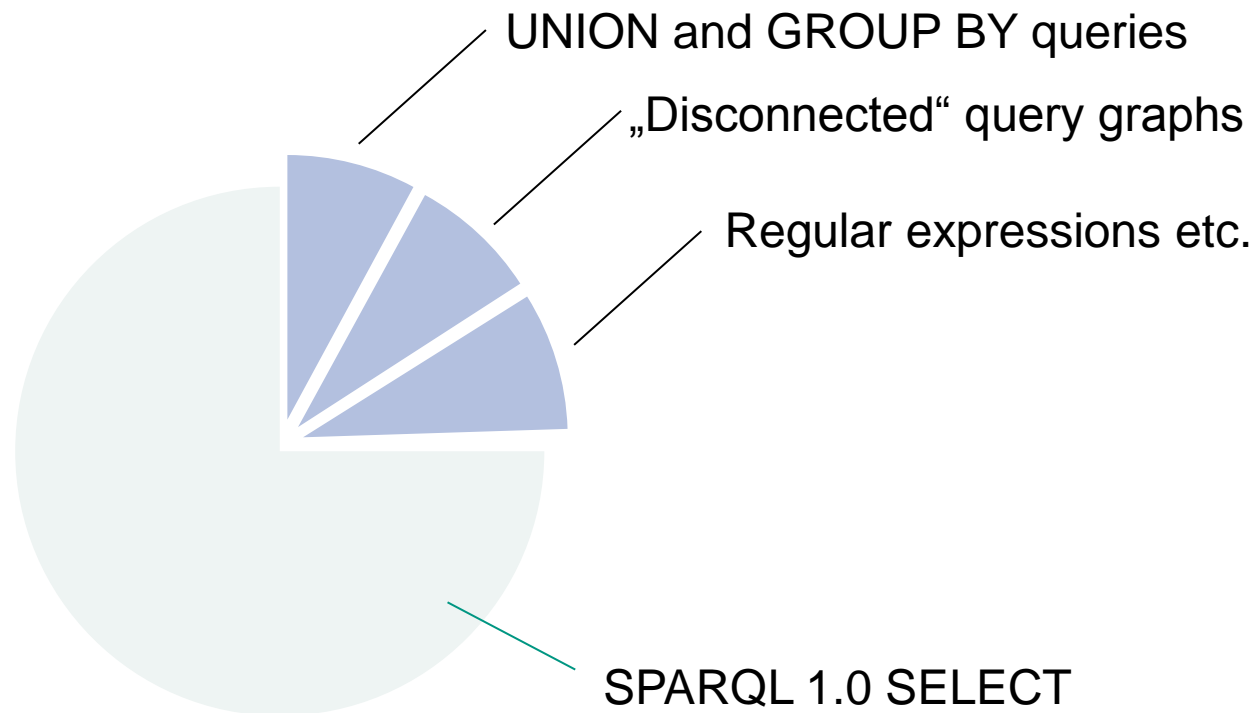
Restrictions



Restrictions



Restrictions

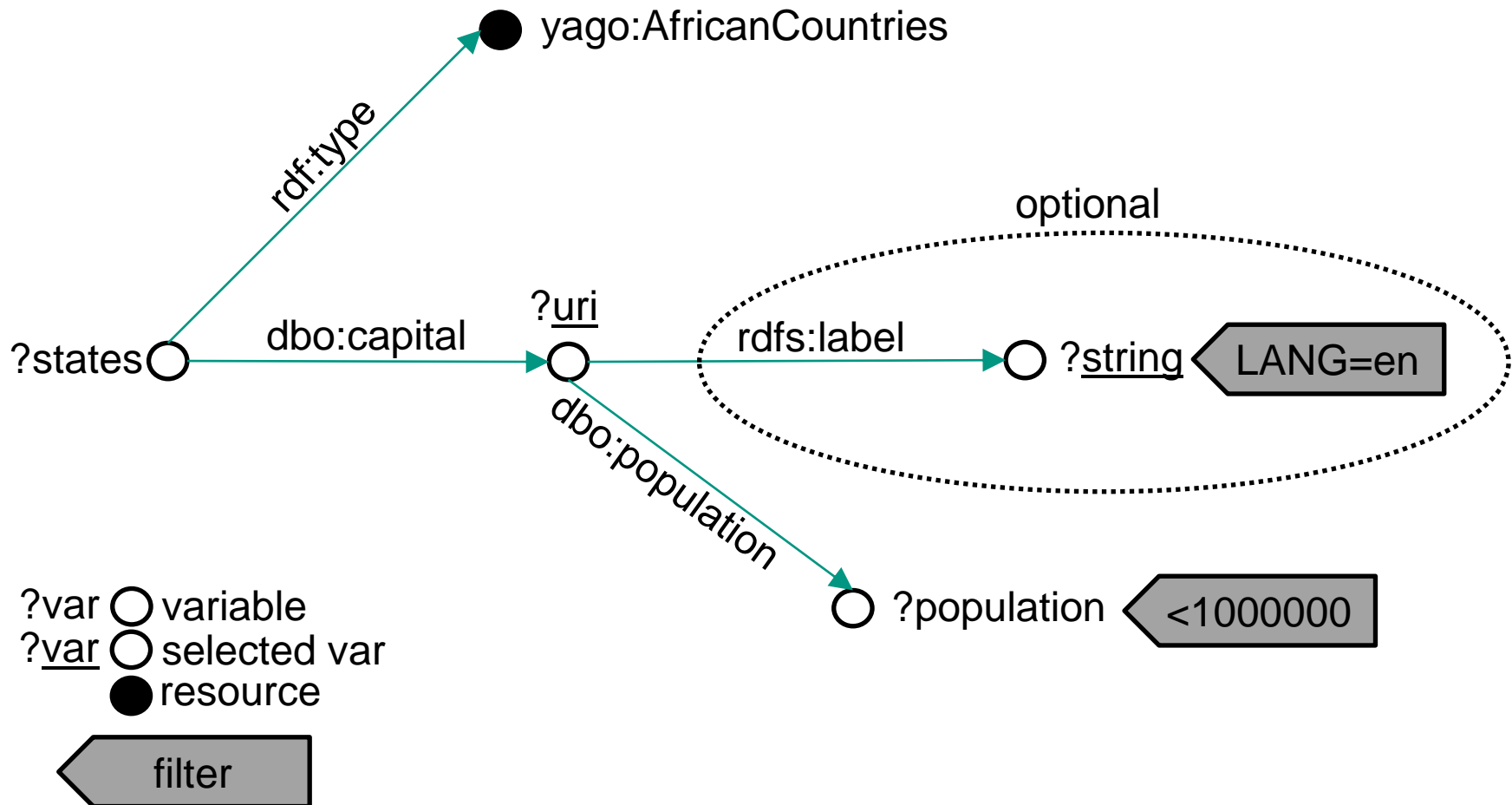


Example – SPARQL query

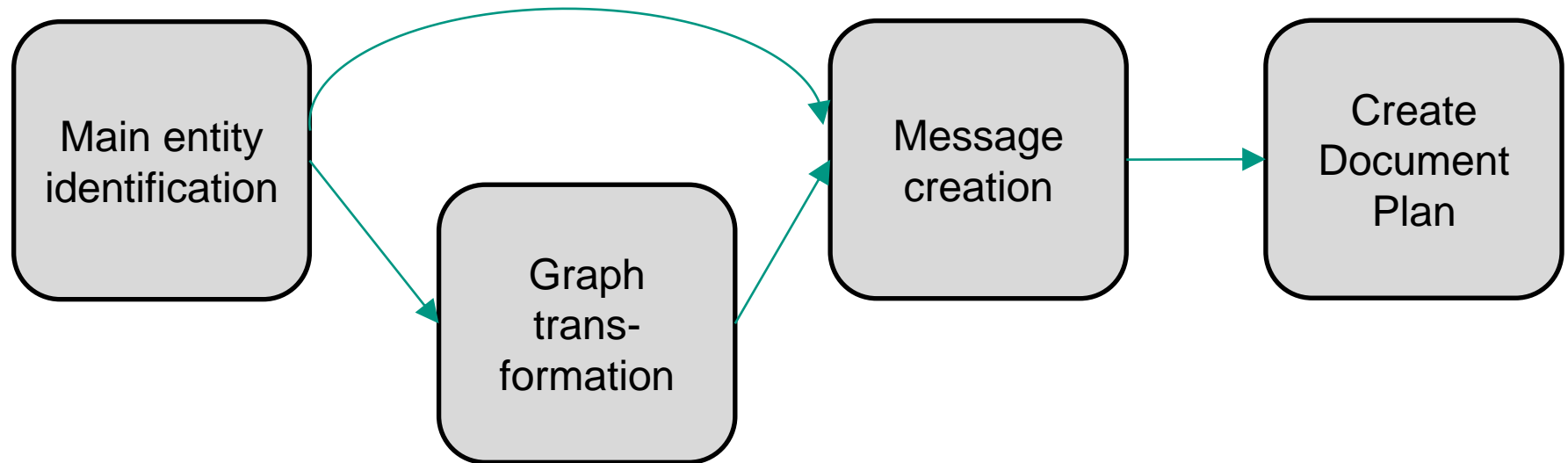
```
01 PREFIX dbo: <http://dbpedia.org/ontology/>
02 PREFIX yago: <http://dbpedia.org/class/yago/>
03 PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
04 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
05 PREFIX dbp: <http://dbpedia.org/property/>

06 SELECT DISTINCT ?uri ?string
07 WHERE {
08     ?states rdf:type yago:AfricanCountries .
09     ?states dbo:capital ?uri .
10     ?uri dbp:population ?population .
11     FILTER ( ?population < 1000000 ) .
12     OPTIONAL { ?uri rdfs:label ?string. FILTER (lang(?string) = 'en') }
13 }
```

Example query – graph representation

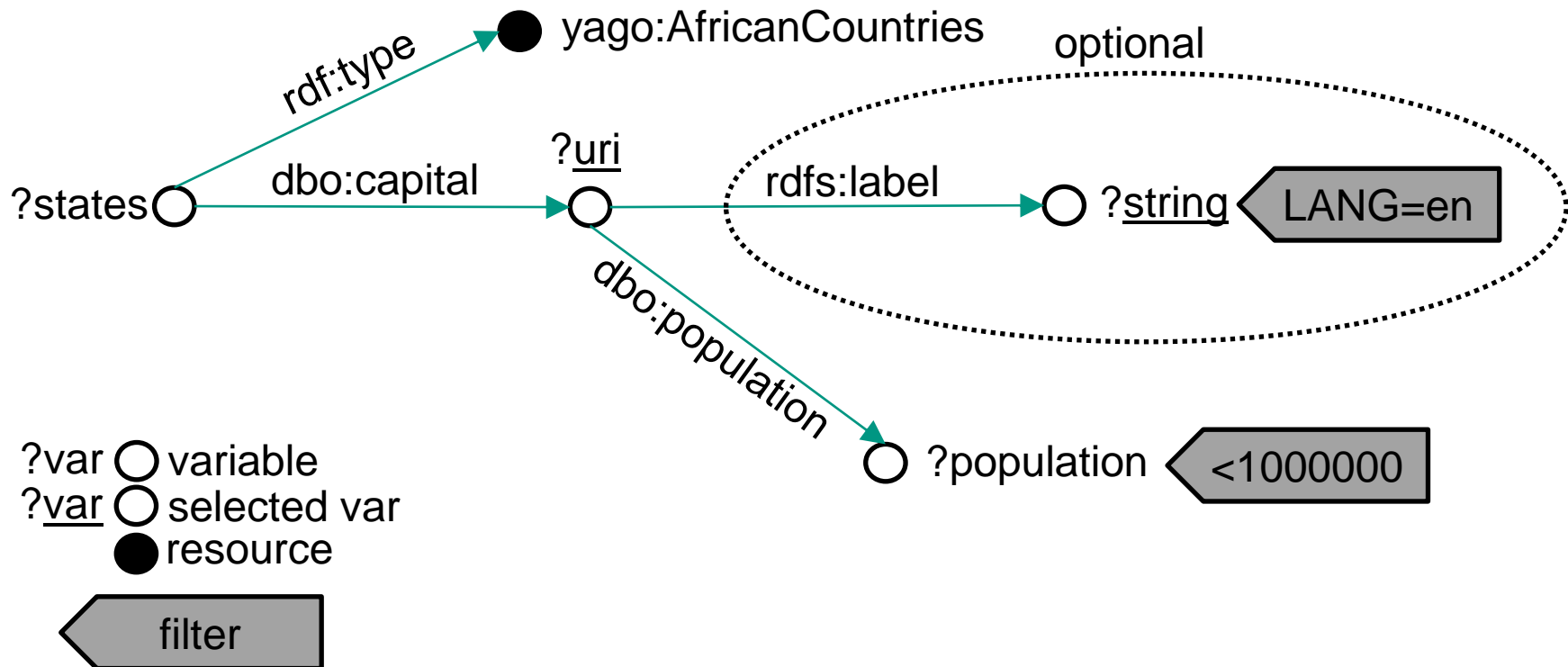


Document structuring – 4 Steps



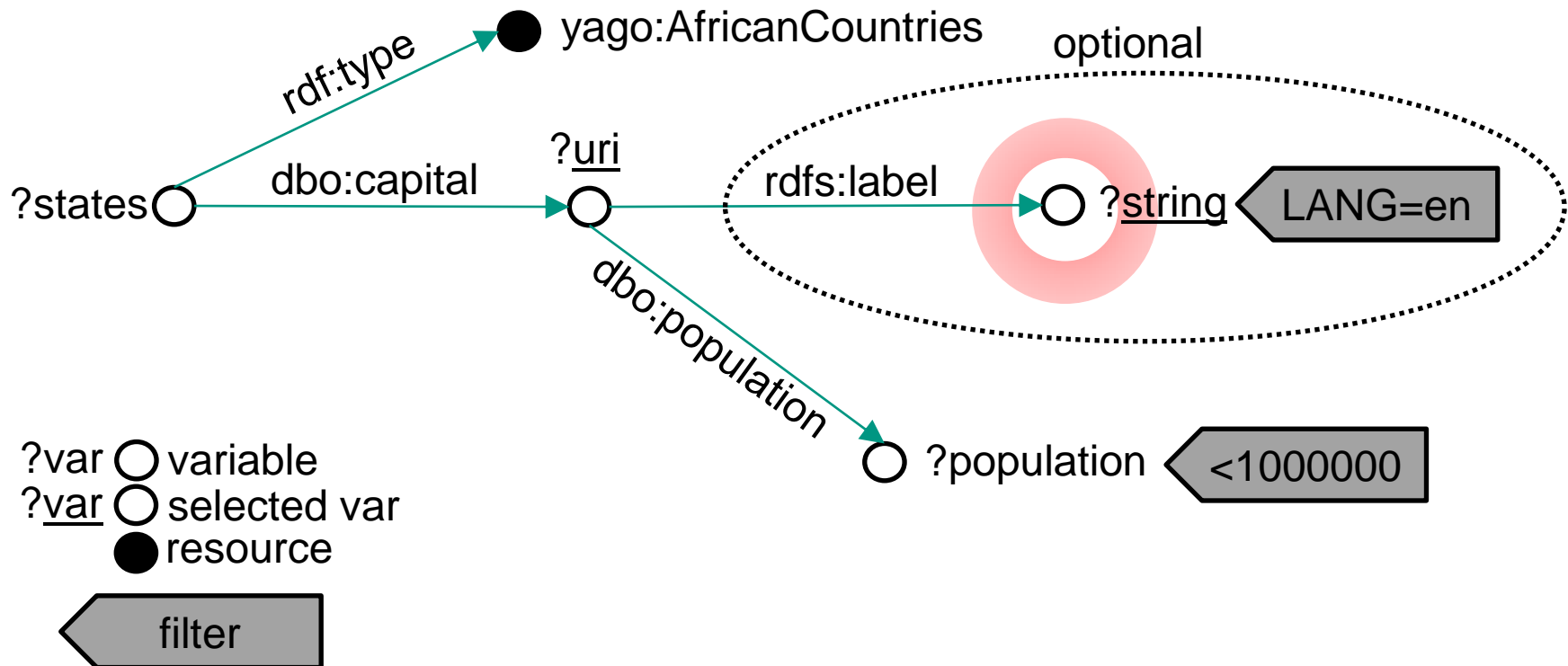
Example – identify main entity

Select a variable that is verbalized as subject



Example – identify main entity

Select a variable that is verbalized as subject

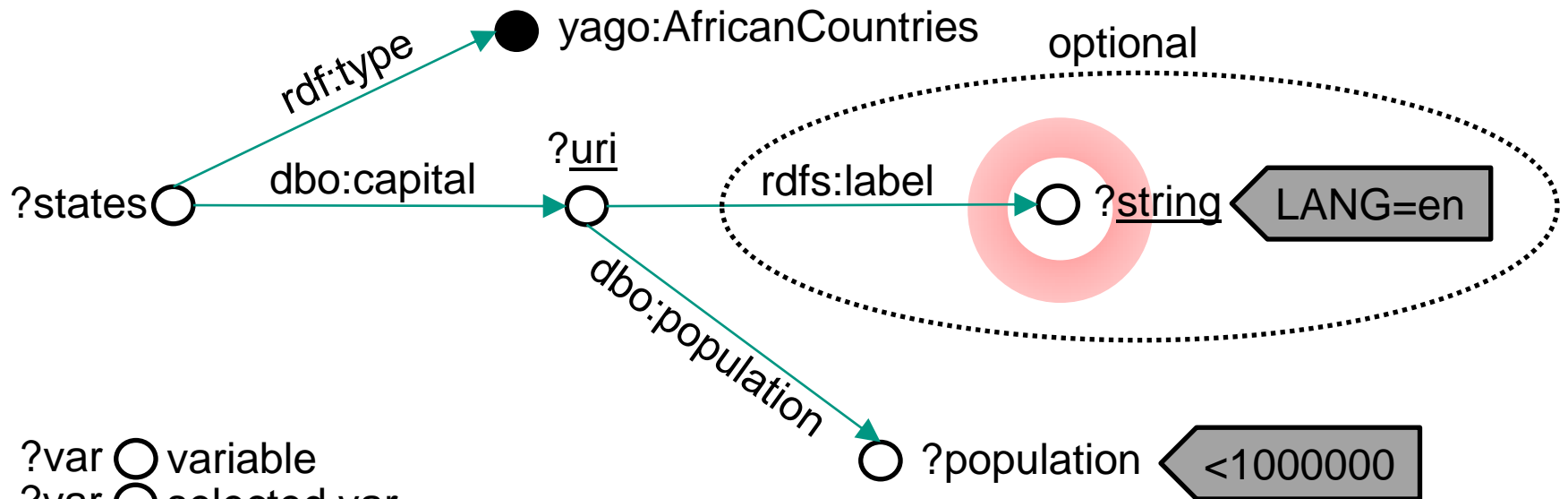


Example – identify main entity

Select a variable that is verbalized as subject

?string

Labels if available of capitals of African countries ...
Bad: subject is optional.



?var ○ variable
 ?var ○ selected var
 ● resource

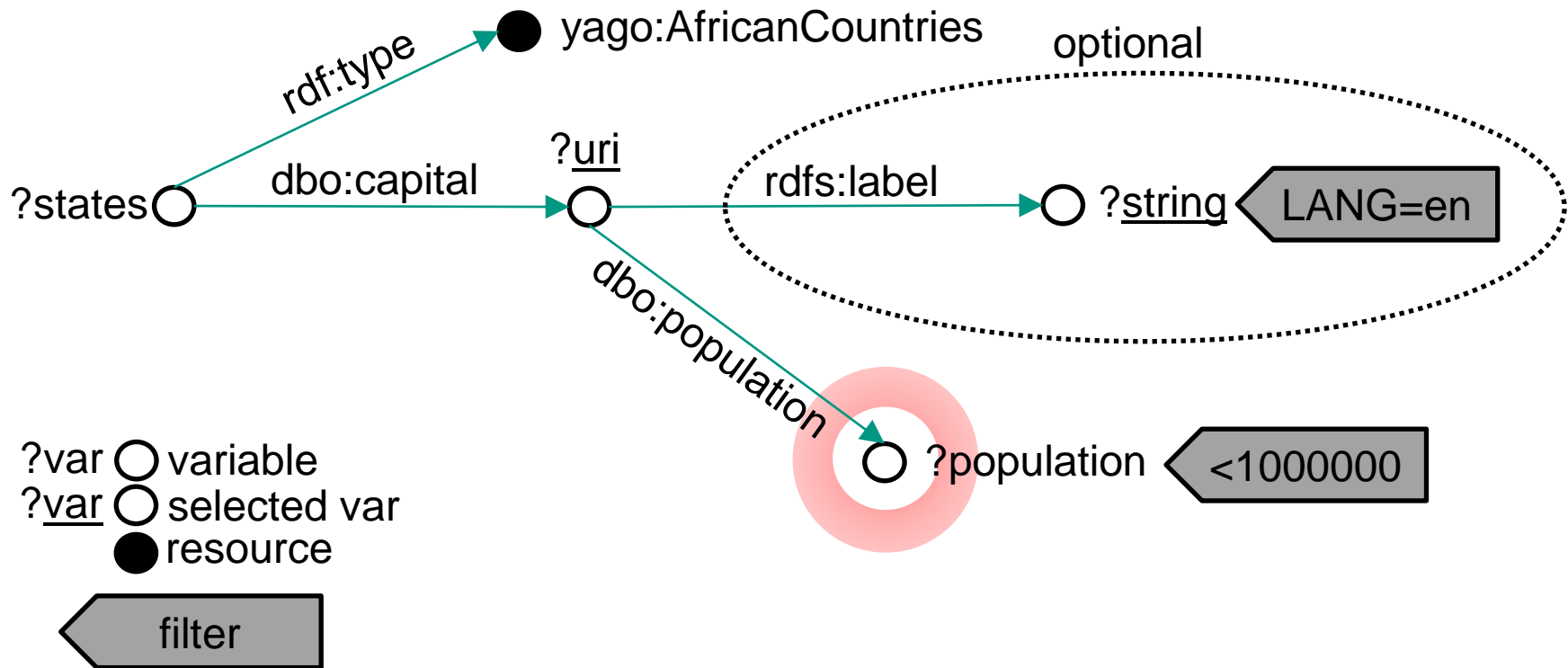
filter

Example – identify main entity

Select a variable that is verbalized as subject

?population
 lation

Population < 10⁶ of capitals of African countries ...
Bad: variable is not selected.

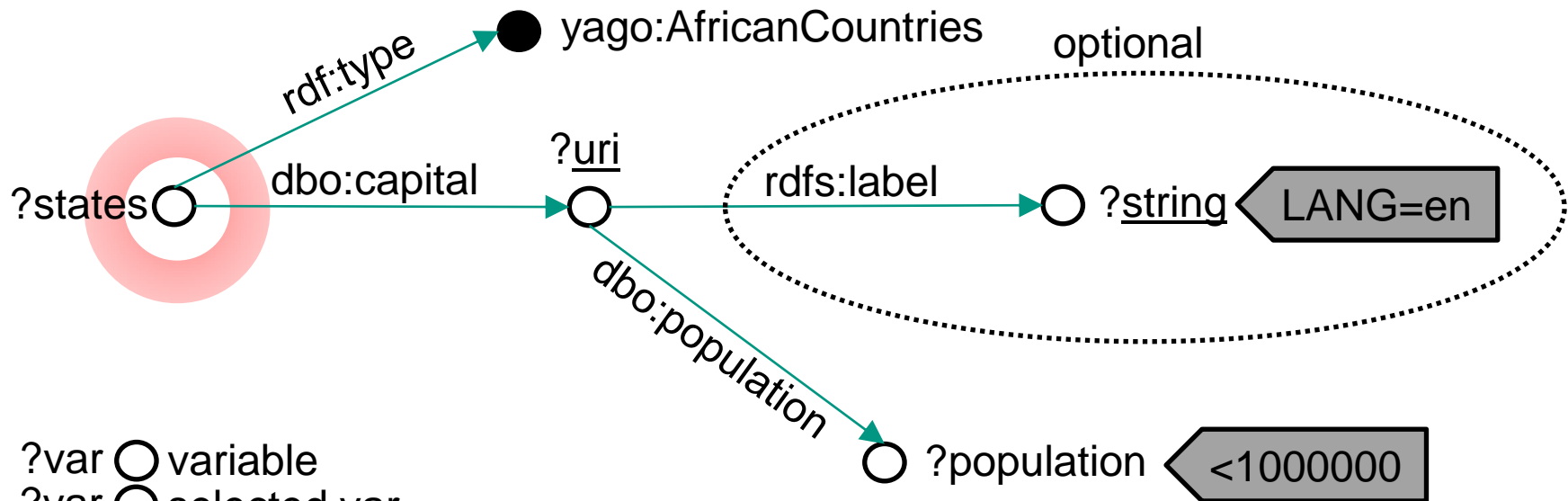


Example – identify main entity

Select a variable that is verbalized as subject

?states

African countries having capitals that have populations < 10⁶ ...
Bad: variable is not selected.



?var ○ variable
 ?var ○ selected var
 ● resource

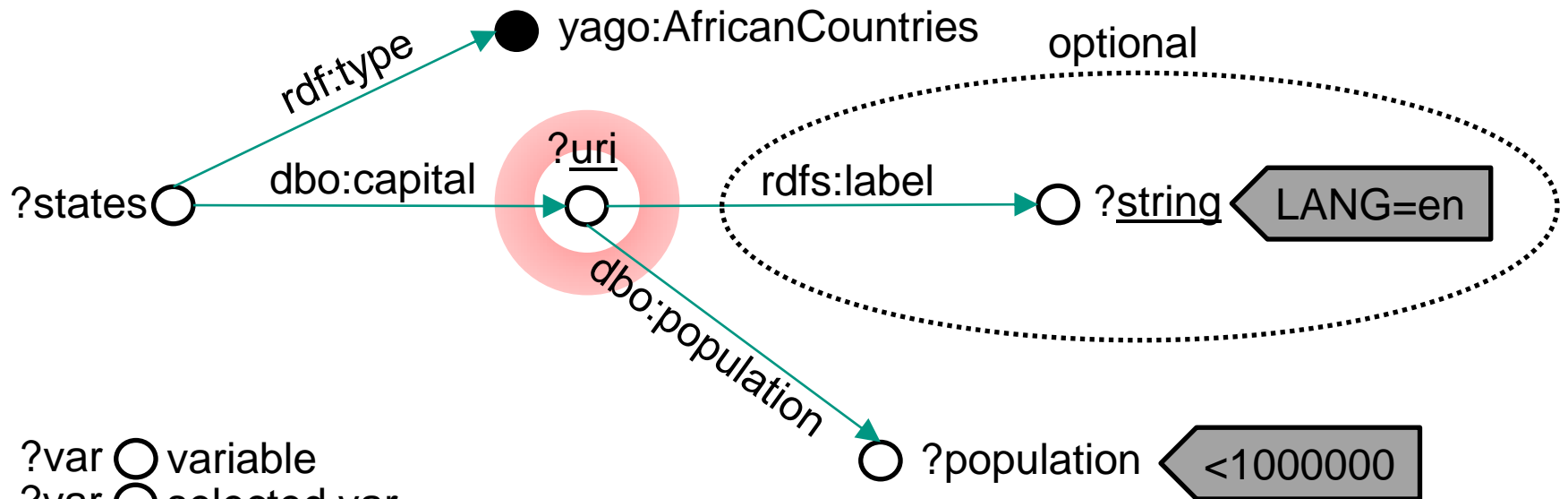
filter

Example – identify main entity

Select a variable that is verbalized as subject

?uri

Capitals of African countries having population < 10⁶ ...
Good: Label for main entity is requested.



?var ○ variable
 ?var ○ selected var
 ● resource

filter

Graph transformation

Idea: Reduce the set of message types
to simplify verbalization

- Main entity is transformed into root node
- Reversal of some edges necessary

Graph transformation

Idea: Reduce the set of message types to simplify verbalization

- Main entity is transformed into root node
- Reversal of some edges necessary

Algorithm 2 Graph transformation

$P \leftarrow \emptyset, F \leftarrow \{(s, p, o) \in Q \mid s = m\}, T \leftarrow \{(s, p, o) \in Q \mid o = m\}$ *(init)*

while $F \neq \emptyset$ or $T \neq \emptyset$ do

 for all $(s_i, p_i, o_i) \in F$ do

 for all $(s_j, p_j, o_j) \in Q \setminus (P \cup F \cup T)$ do

 if $o_i = s_j$ then

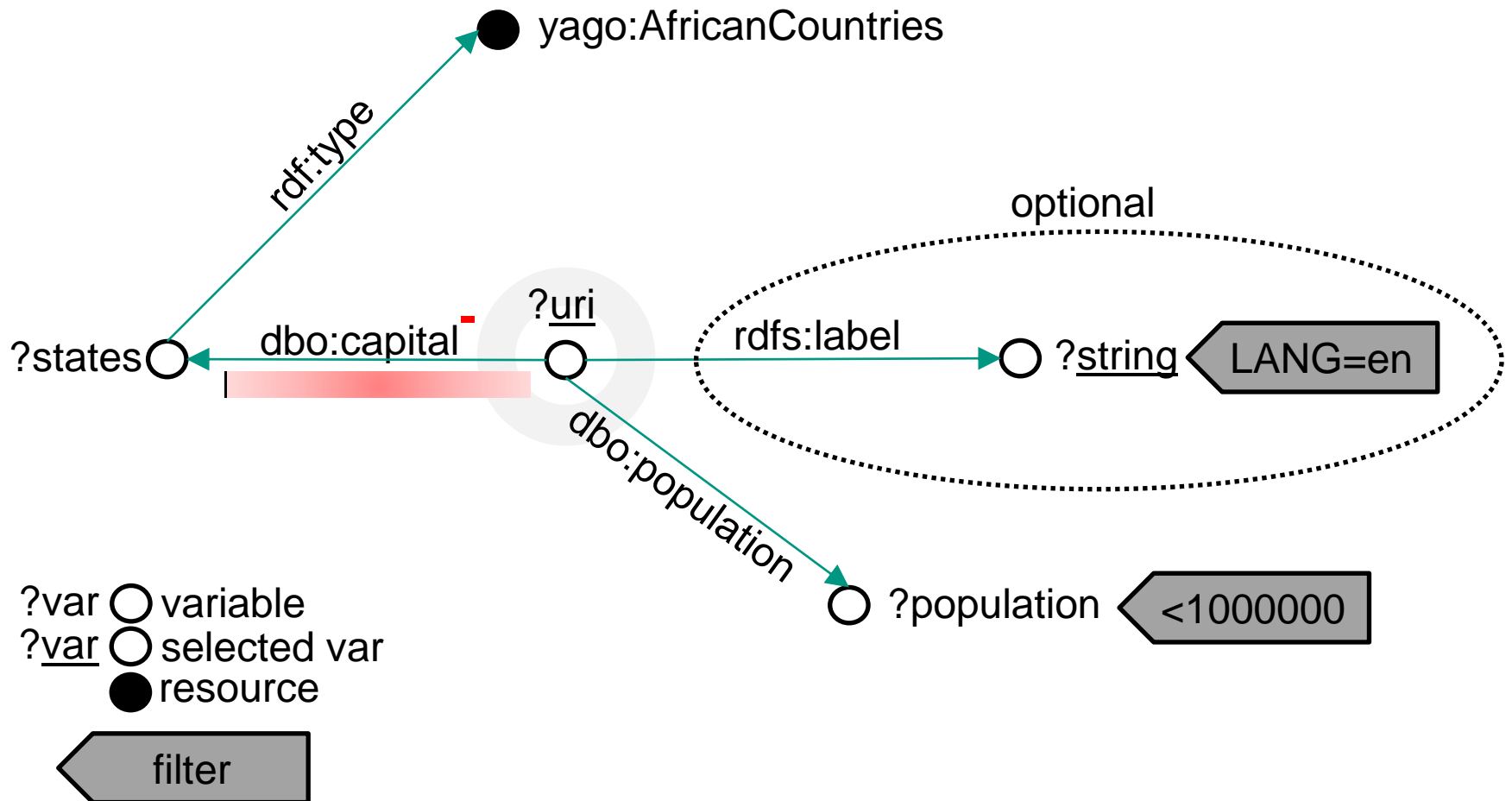
$F \leftarrow F \cup \{(s_j, p_j, o_j)\}$

 else if $o_i = o_j$ then

$T \leftarrow T \cup \{(s_j, p_j, o_j)\}$

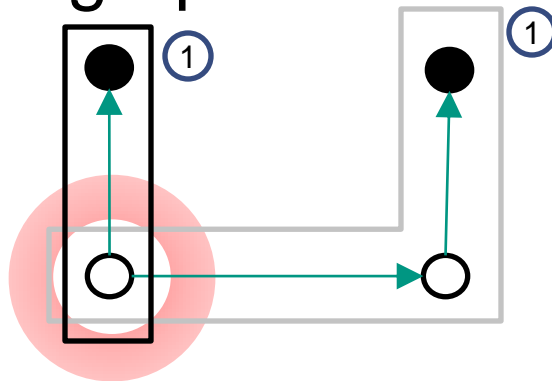


Example – transformed graph



Message creation

- Cut graph into independently verbalizable parts

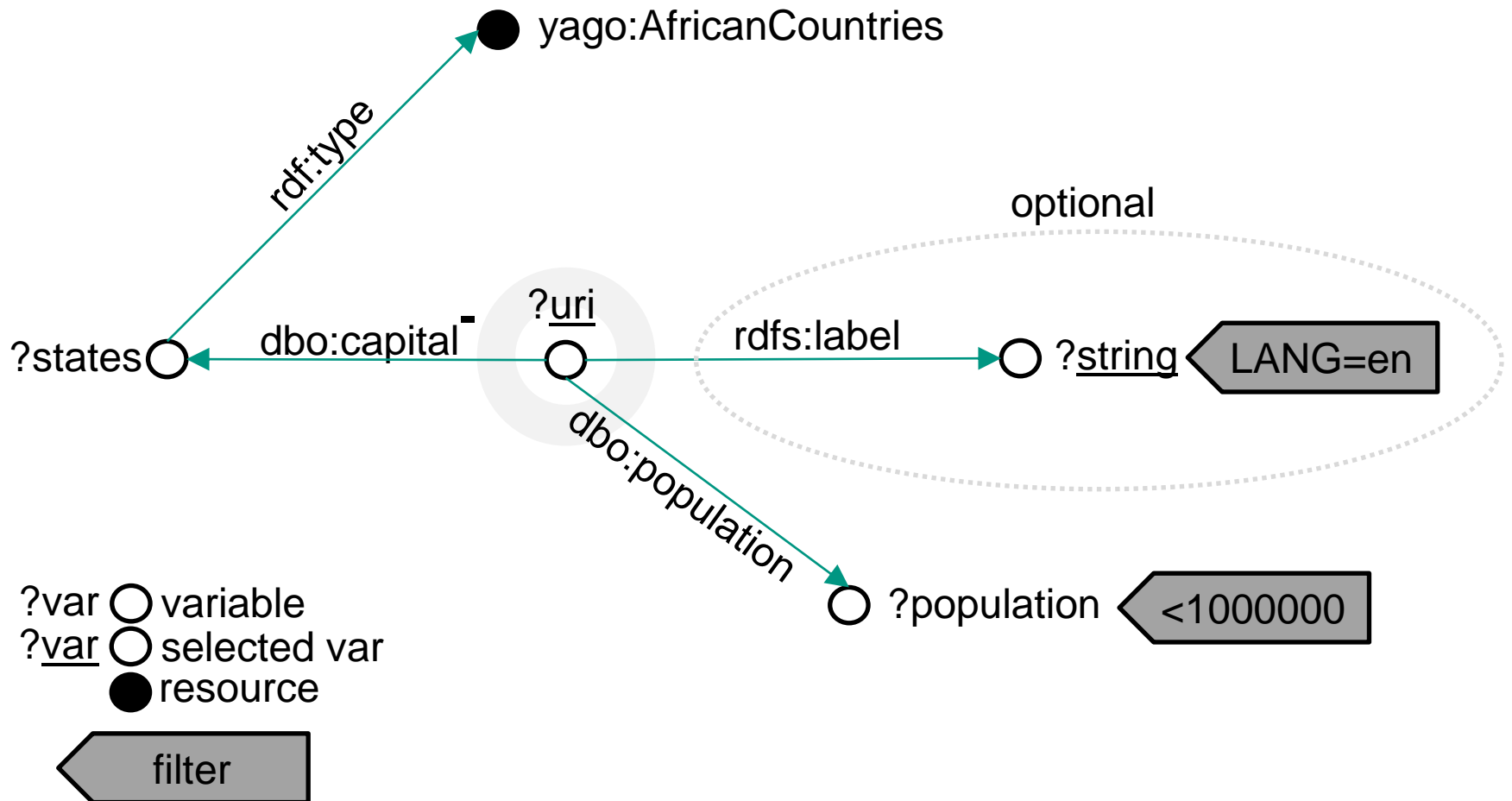


Messages (1-9) represent paths,
message types are path classes

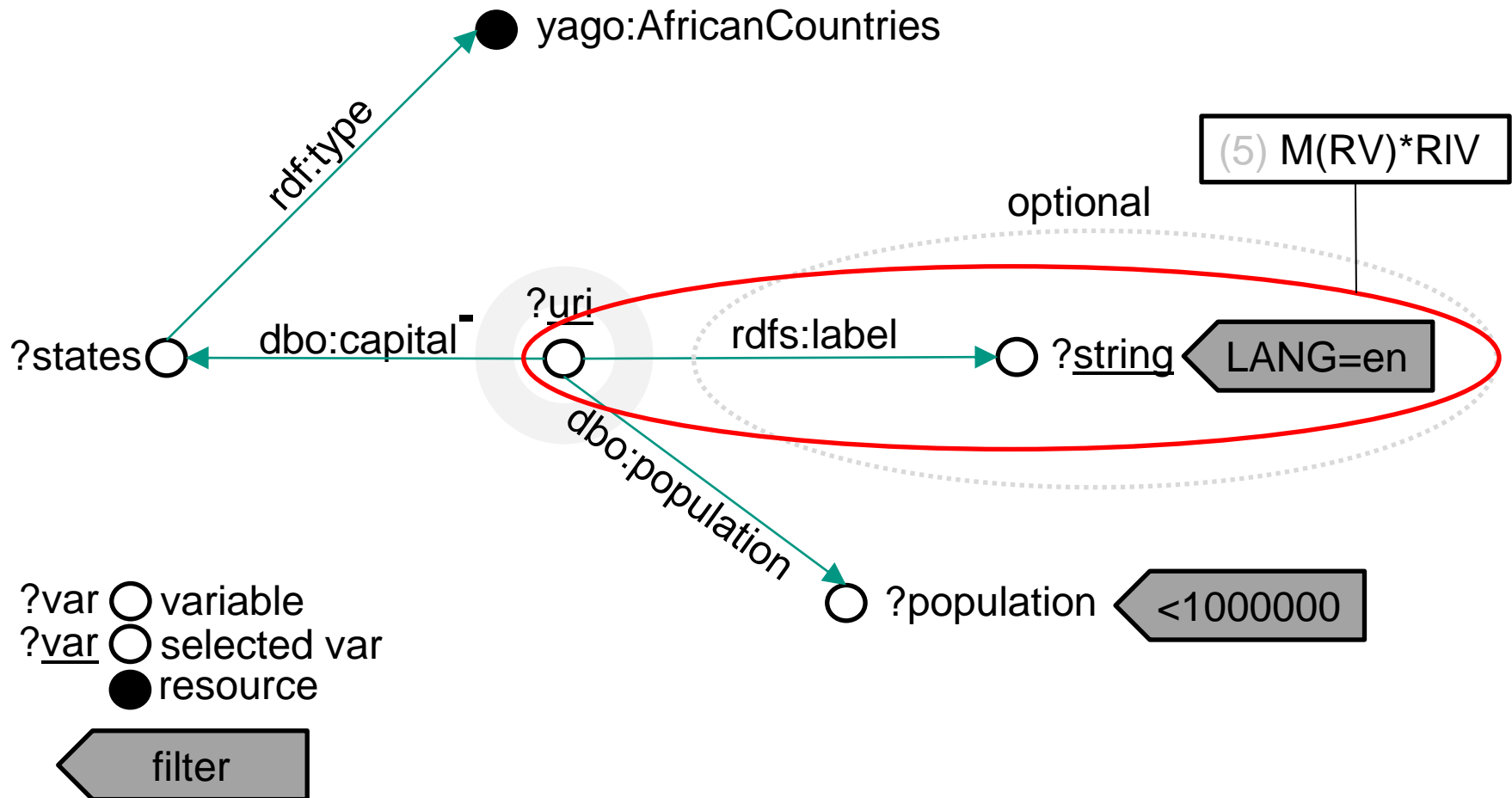
nr name	nr name	nr name
(1) $M(RV)^*RR$	(2) $M(RV)^*RL$	(3) $M(RV)^*RV$
(4) $M(RV)^*R_lR$	(5) $M(RV)^*R_lV$	(6) $M(RV)^*R_lL$
(7) $M(RV)^*R_tR$	(8) $M(RV)^*R_tV$	(9) $M(RV)^*R_tL$
(10) VAR	(11) $ORDERBY$	(12) $LIMIT$
(13) $OFFSET$	(14) $HAVING$	

- Filters are stored in VAR messages

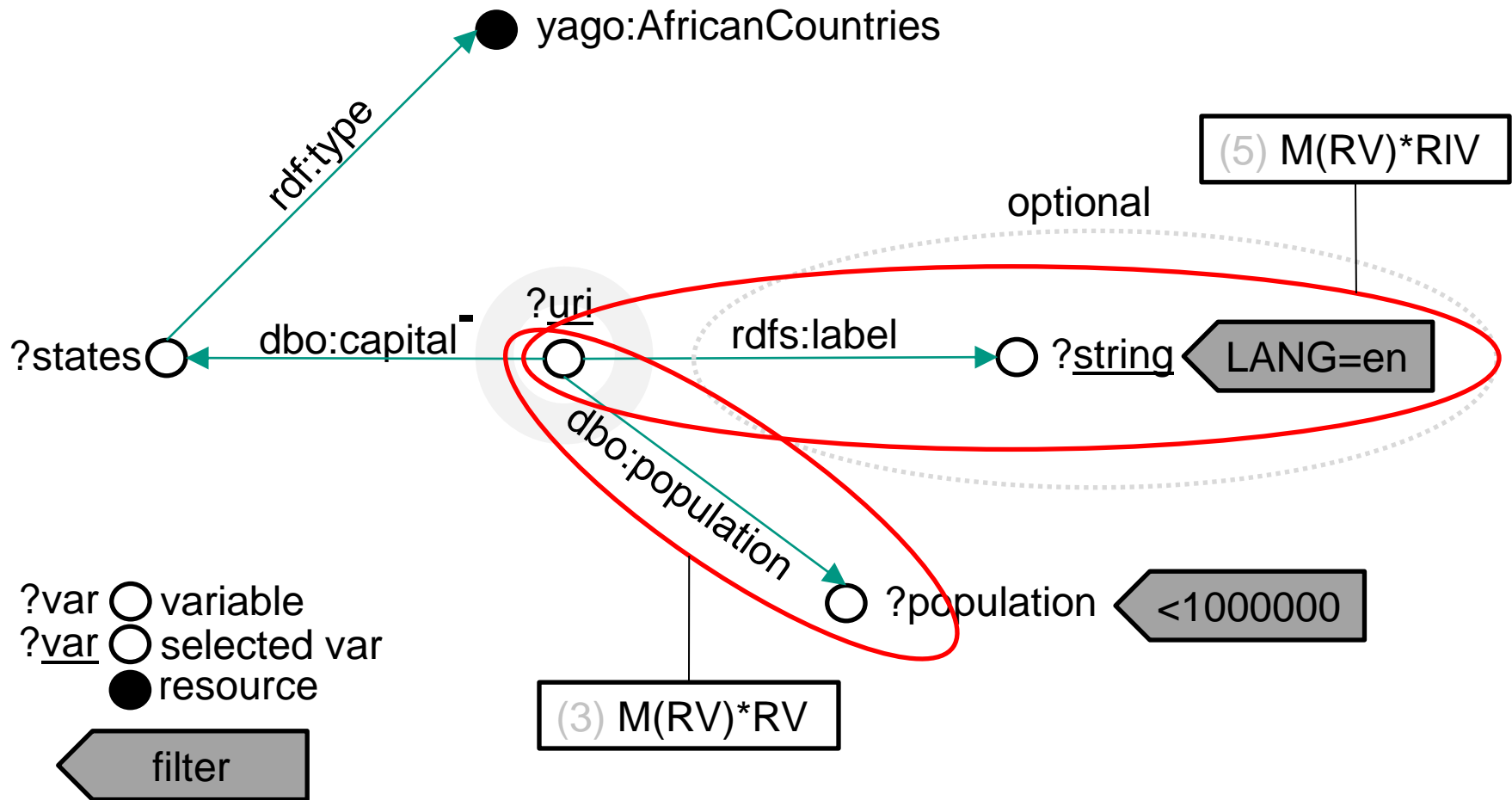
Example – messages



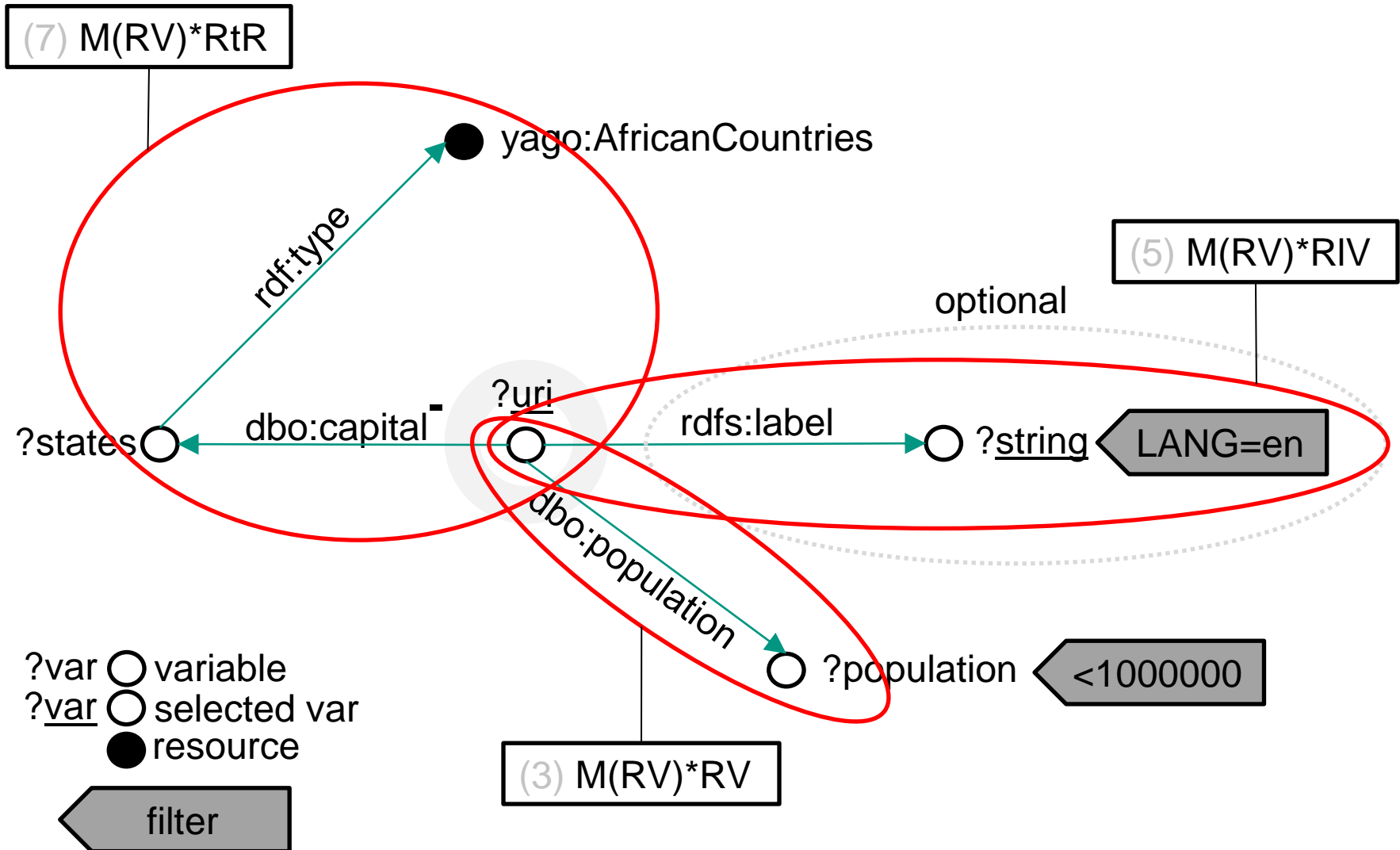
Example – messages



Example – messages



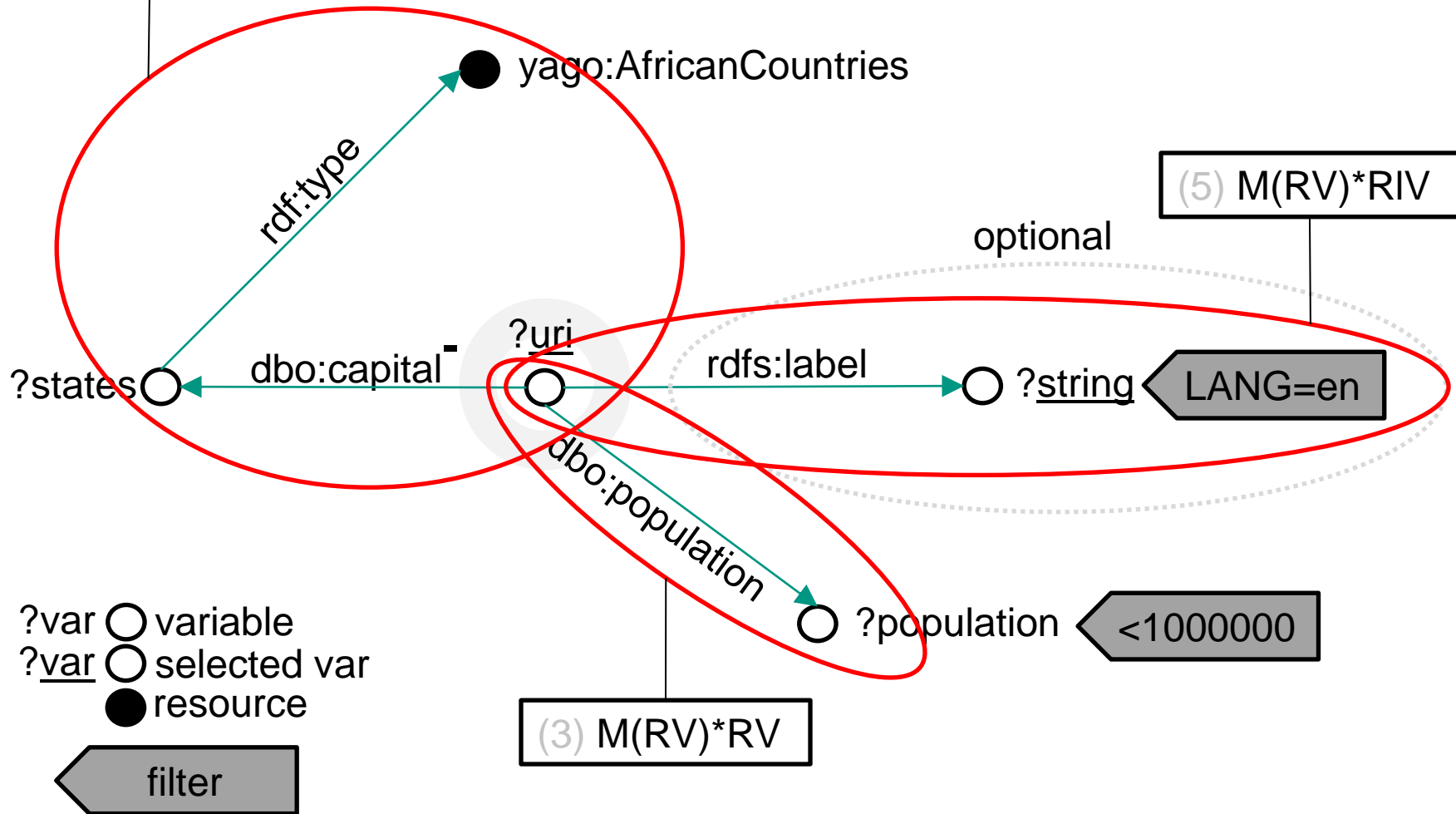
Example – messages



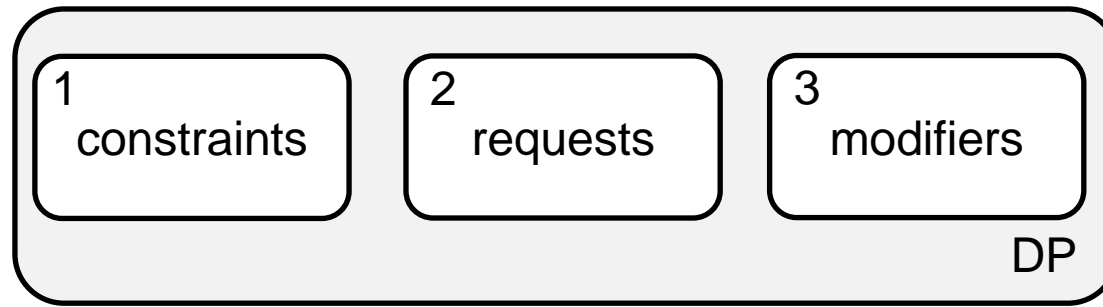
Example – messages

(7) M(RV)*RtR

+4 x (10) VAR



Document Plan

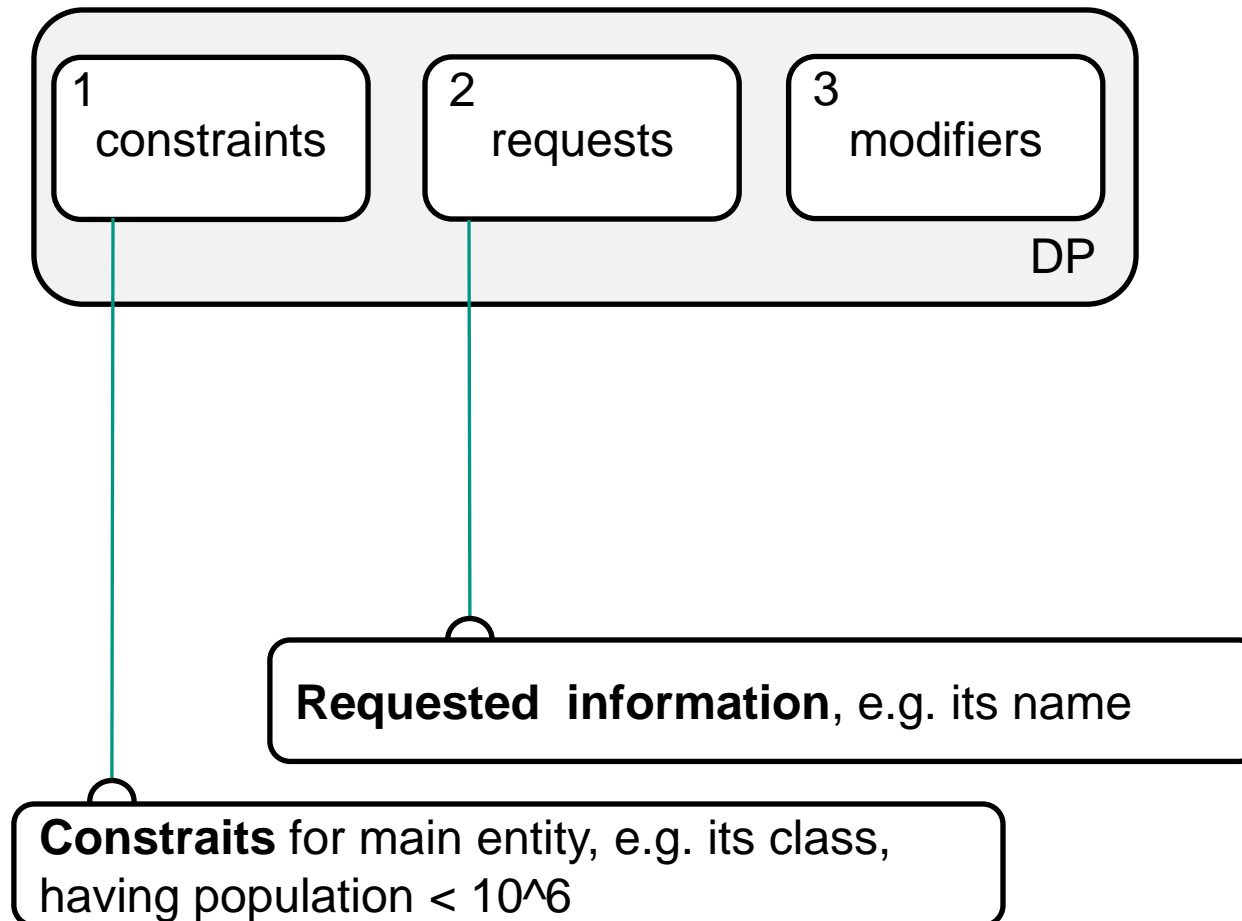


Document Plan

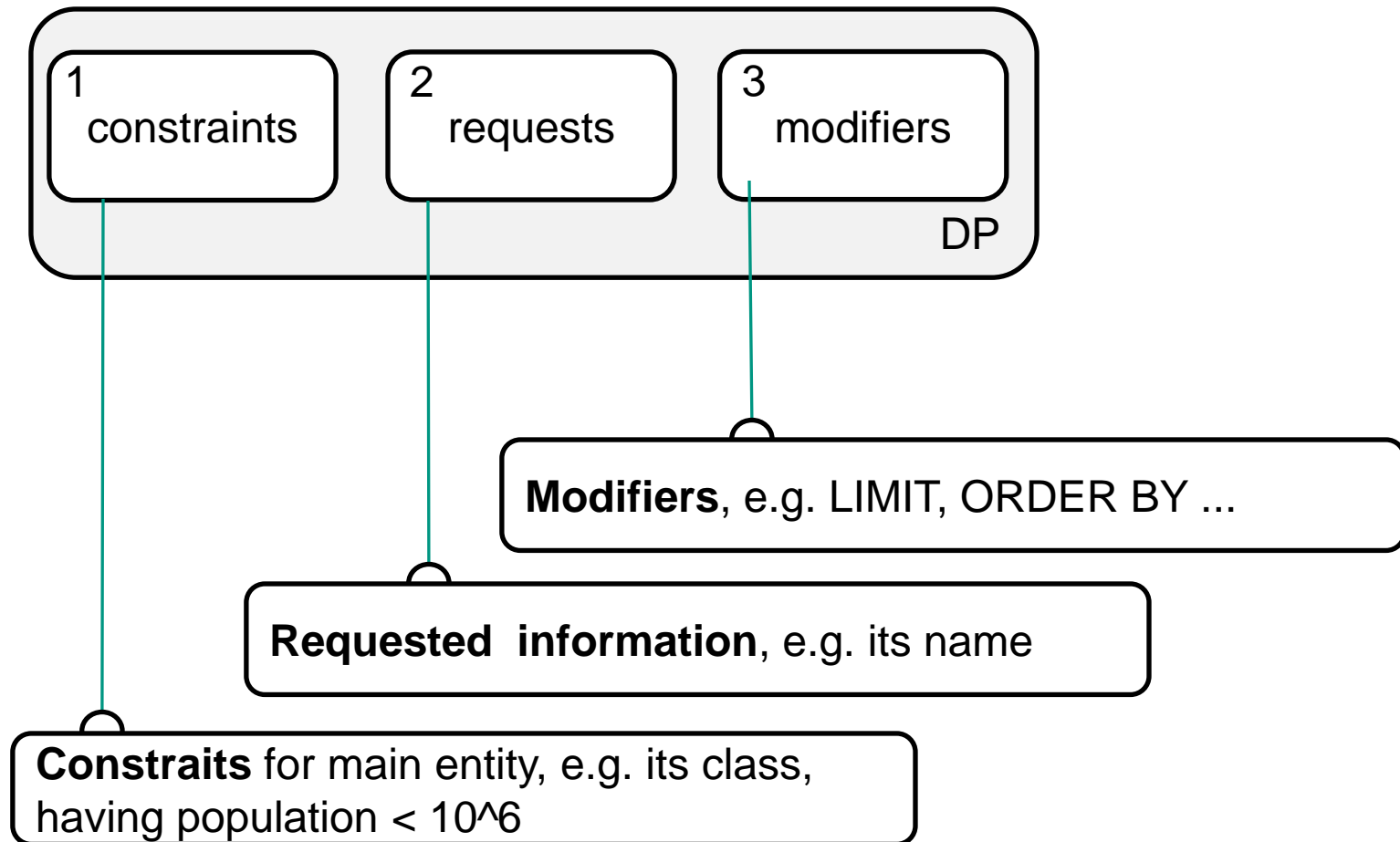


Constraints for main entity, e.g. its class,
having population $< 10^6$

Document Plan

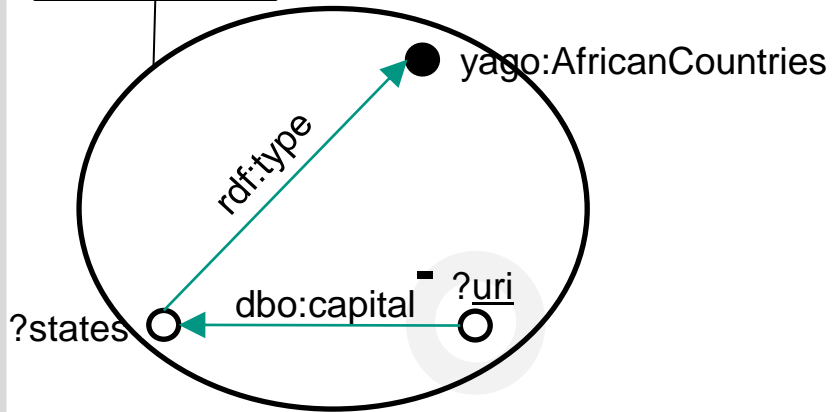


Document Plan



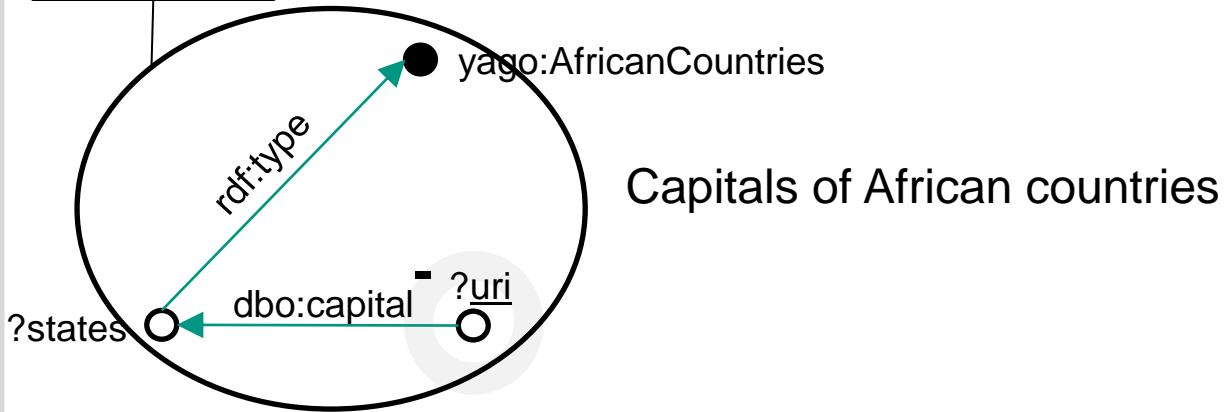
Example - verbalization

$M(RV)^*RtR$ (cons)

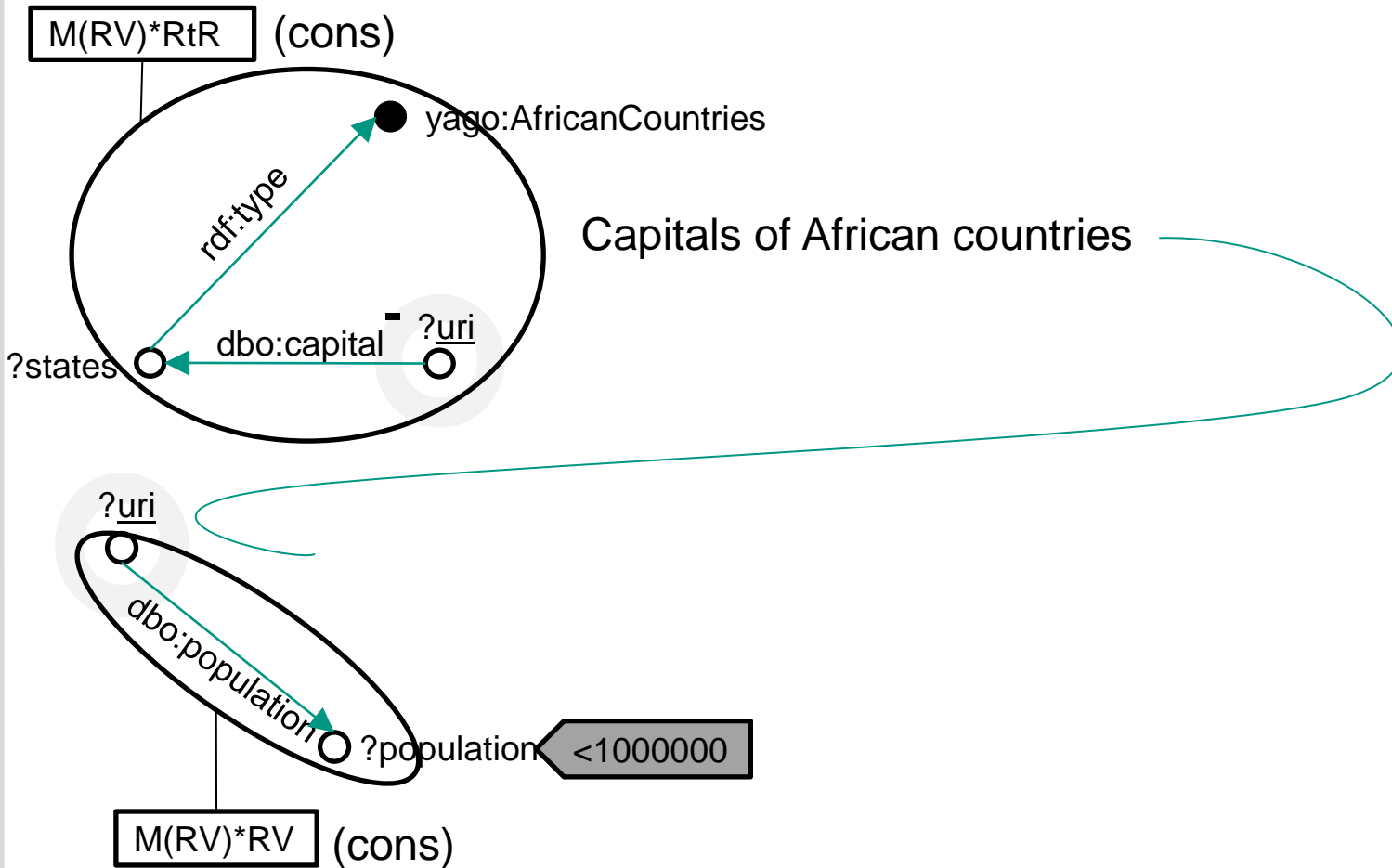


Example - verbalization

$M(RV)^*RtR$ (cons)

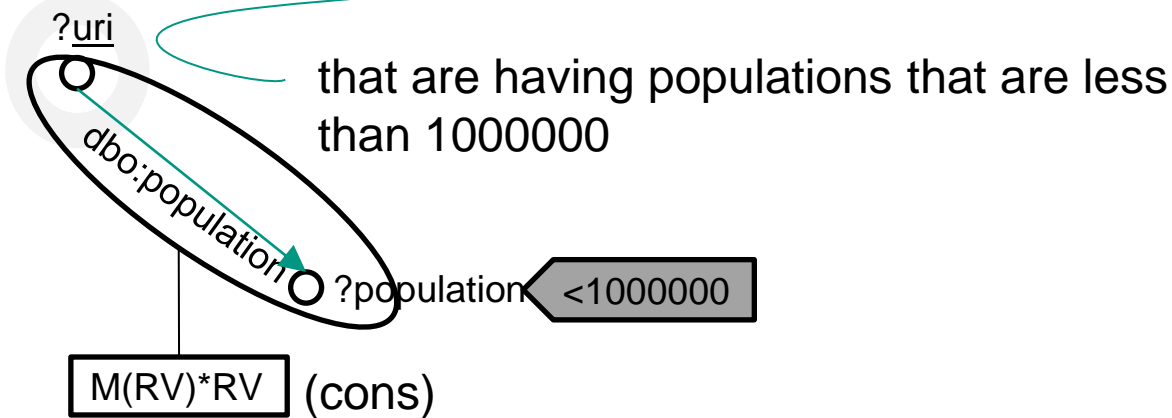
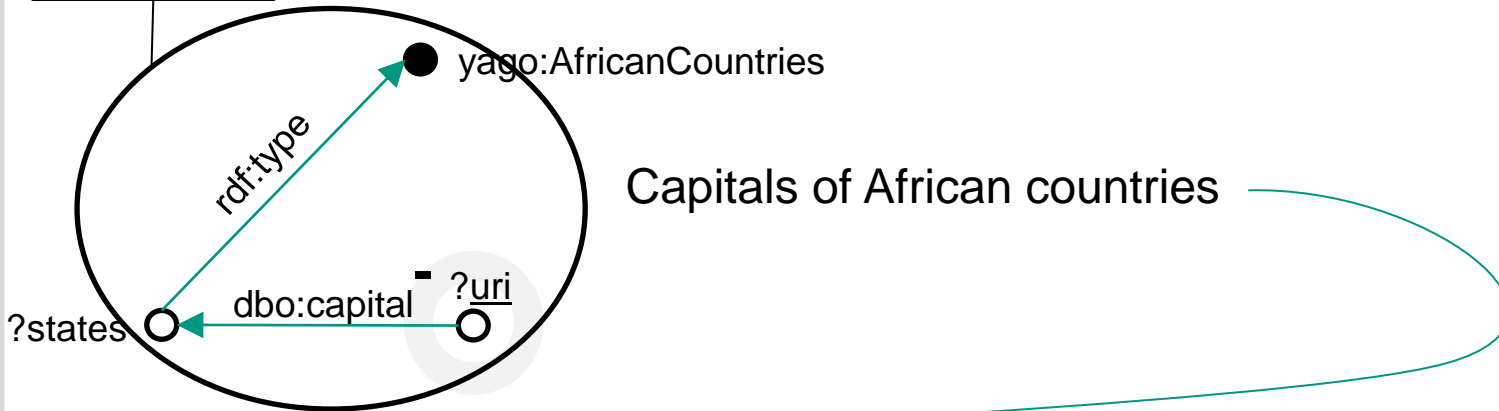


Example - verbalization



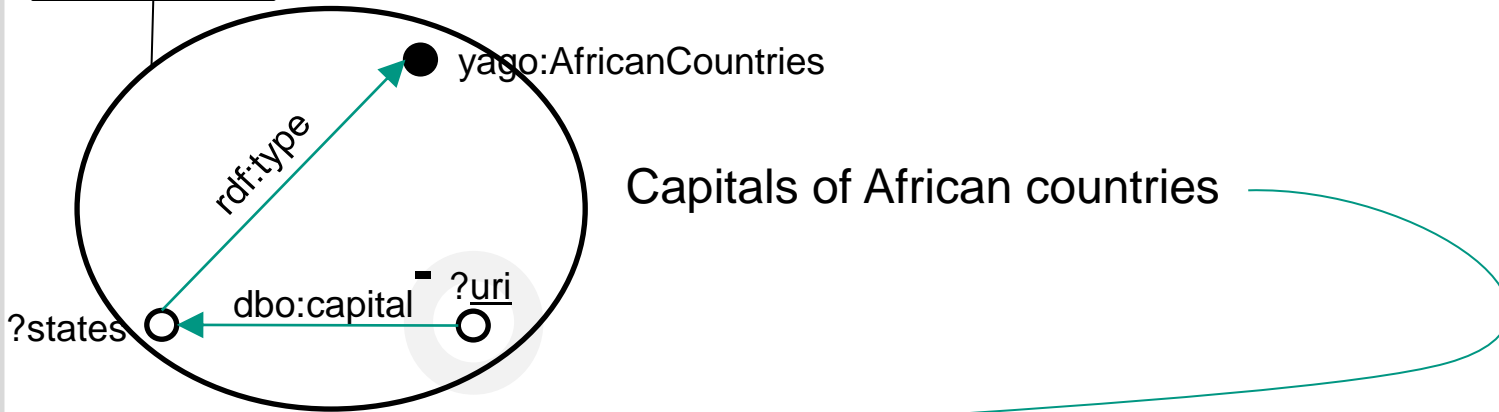
Example - verbalization

$M(RV) * R_tR$ (cons)

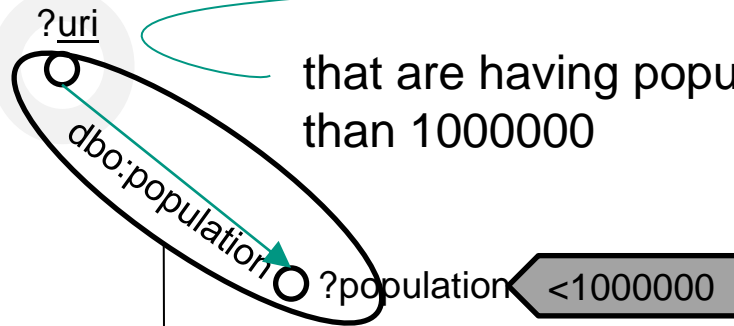


Example - verbalization

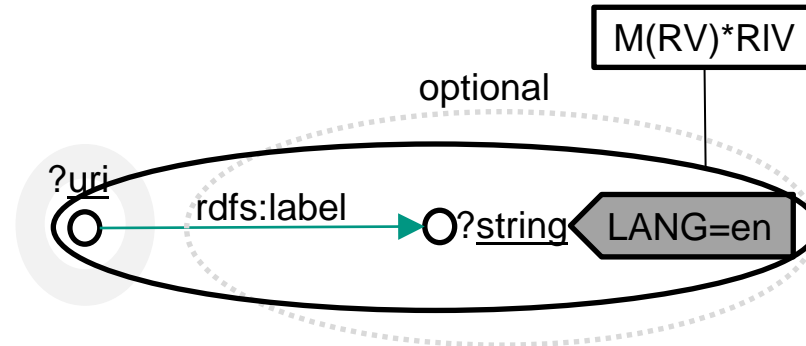
$M(RV)*RtR$ (cons)



that are having populations that are less than 1000000

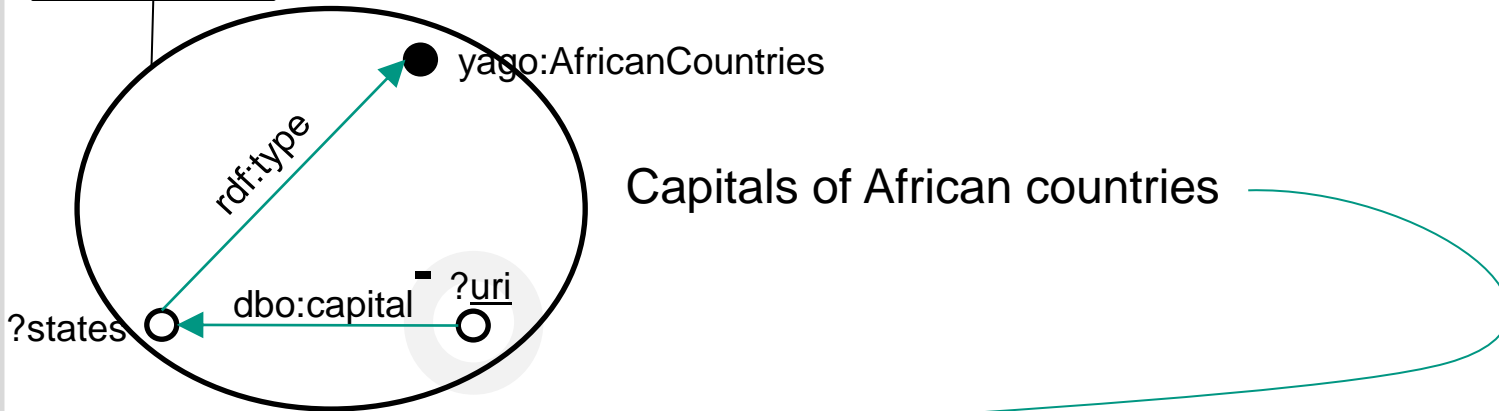


$M(RV)*RIV$ (req)

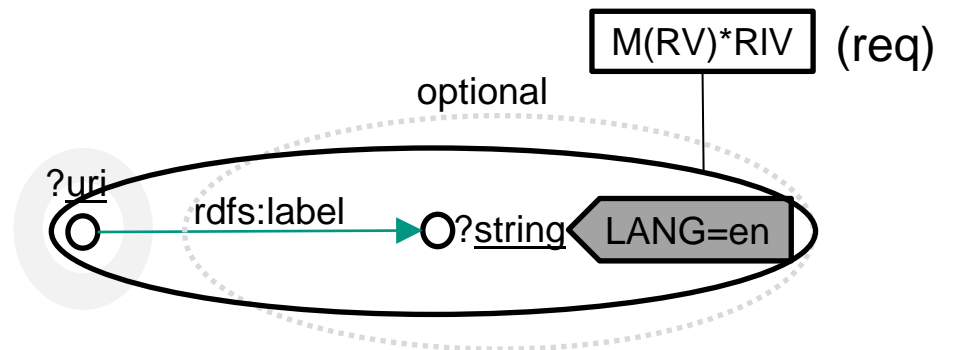
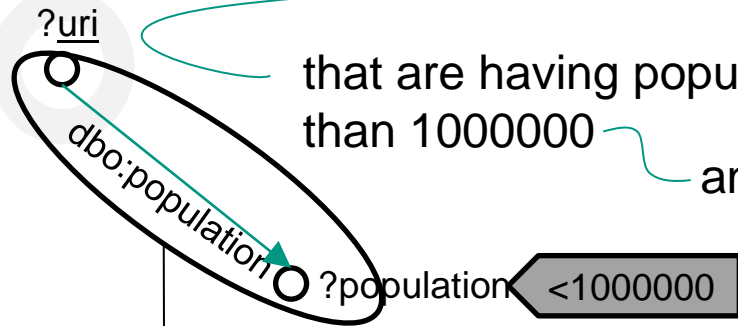


Example - verbalization

$M(RV)*RtR$ (cons)



that are having populations that are less than 1000000 and where available their English labels.



SUMMARY AND FUTURE WORK

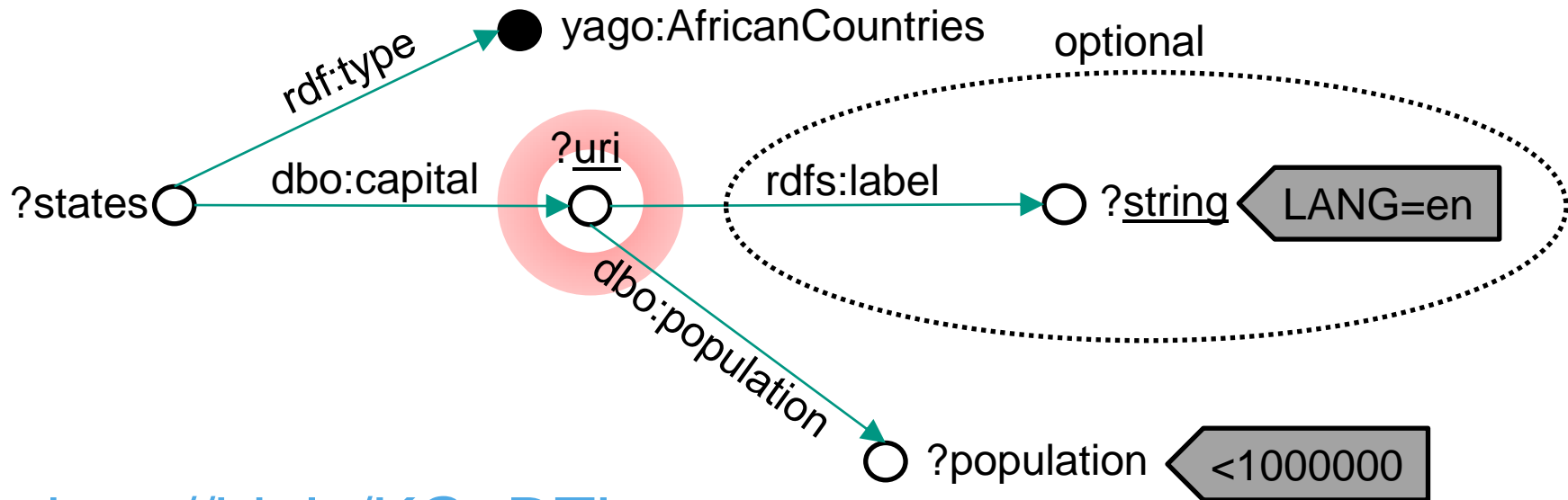
Summary and Future Work

- Summary:
 - Presented an approach for explaining SPARQL SELECT queries in natural language
 - Schema-agnostic

Summary and Future Work

- Summary:
 - Presented an approach for explaining SPARQL SELECT queries in natural language
 - Schema-agnostic

- Directions for future work:
 - Tackle challenges in the two missing pipeline components
 - Exploitation of linguistic features of labels
 - Evaluation



<http://bit.ly/KGuDTL>

<http://km.aifb.kit.edu/projects/spartiqulator/>

?QUESTIONS

The work presented here is supported by the European Union's 7th Framework Programme (FP7/2007-2013) under Grant Agreement 257790.



REFERENCES

References

S. Shekarpour, S. Auer, A.-C. Ngonga Ngomo, D. Gerber, S. Hellmann, and C. Stadler. Keyword-driven SPARQL Query Generation Leveraging Background Knowledge. In International Conference on Web Intelligence, 2011.

E. Reiter and R. Dale. Building Natural Language Generation Systems. Natural Language Processing. Cambridge University Press, 2000.

P. Haase, D. M. Herzig, M. Musen, and D. T. Tran. Semantic Wiki Search. In L. A. P. et al., editor, 6th Annual European Semantic Web Conference, ESWC2009, Heraklion, Crete, Greece, volume 5554 of LNCS, pages 445-460. Springer Verlag, Juni 2009.

QALD 2011: <http://www.sc.cit-ec.uni-bielefeld.de/qald-1>