DBCLS

Natural Language Interfaces for SPARQL endpoints
- Hands-on tutorial on LODQA -

Jin-Dong Kim (DBCLS)

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

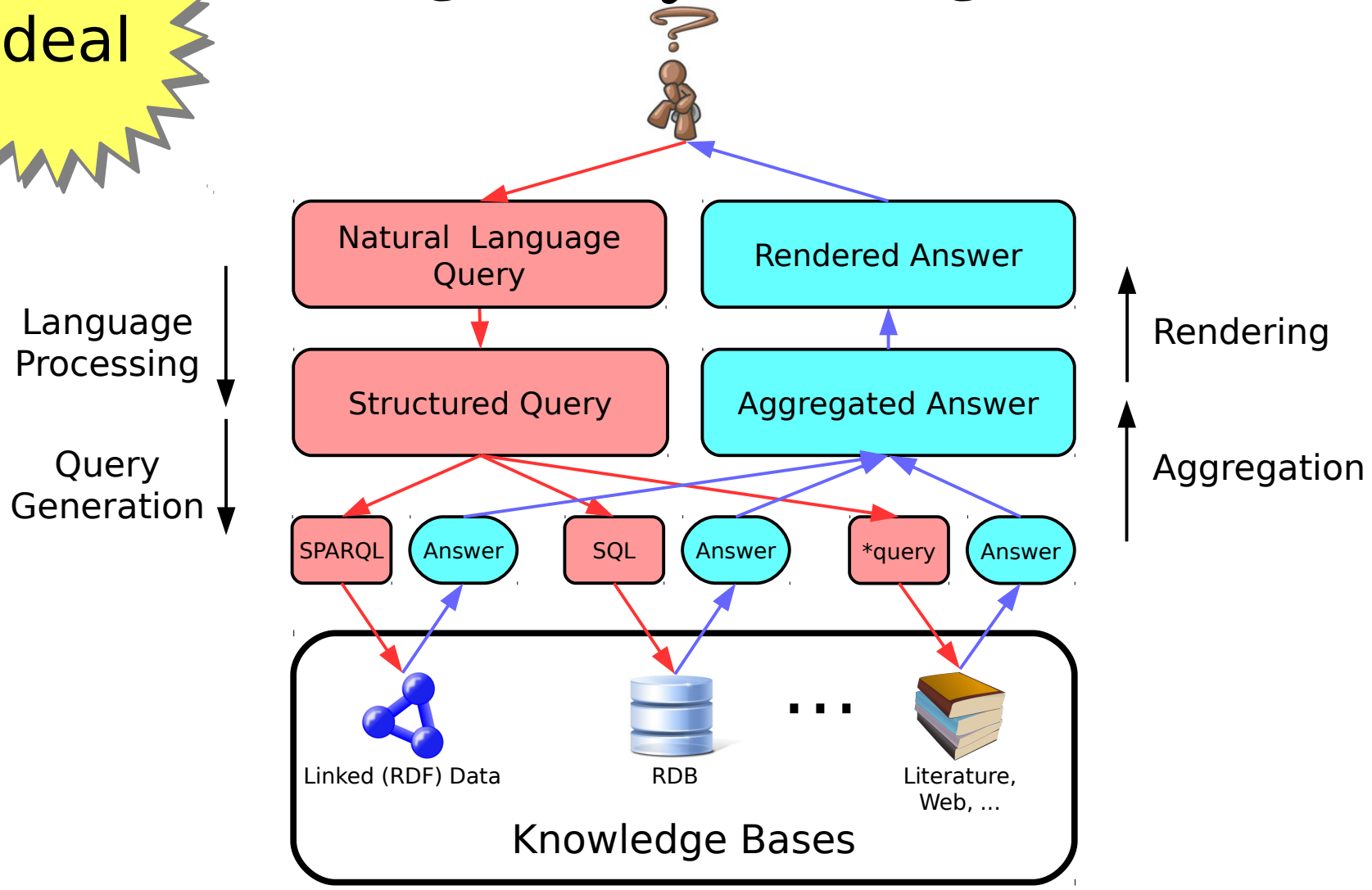
Agenda

- Intro to NLI SPARQL
- LODQA intro
- LODQA hands-on
- Related works

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

NLQA (Hybrid QA)

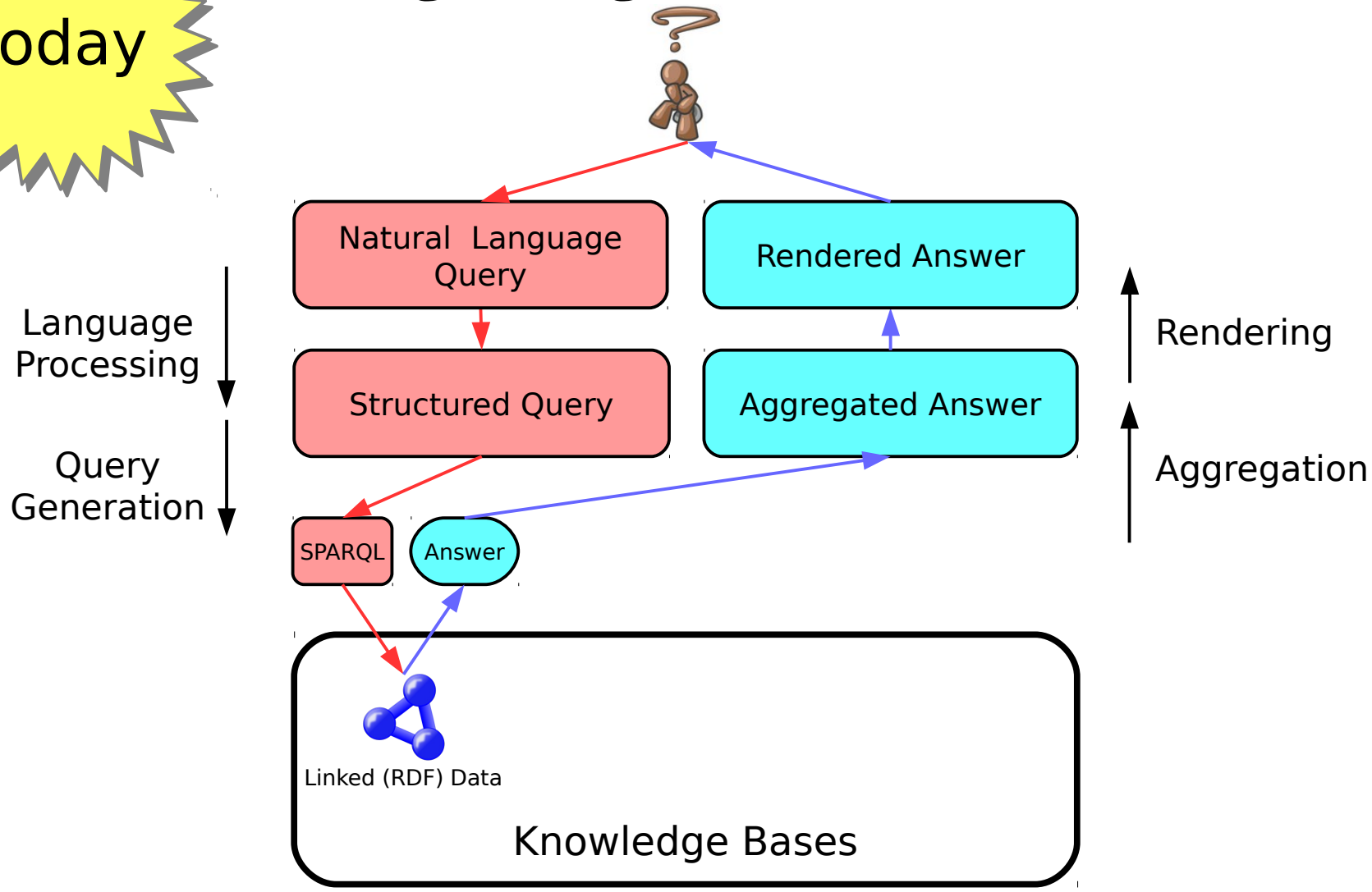
Ideal



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

today

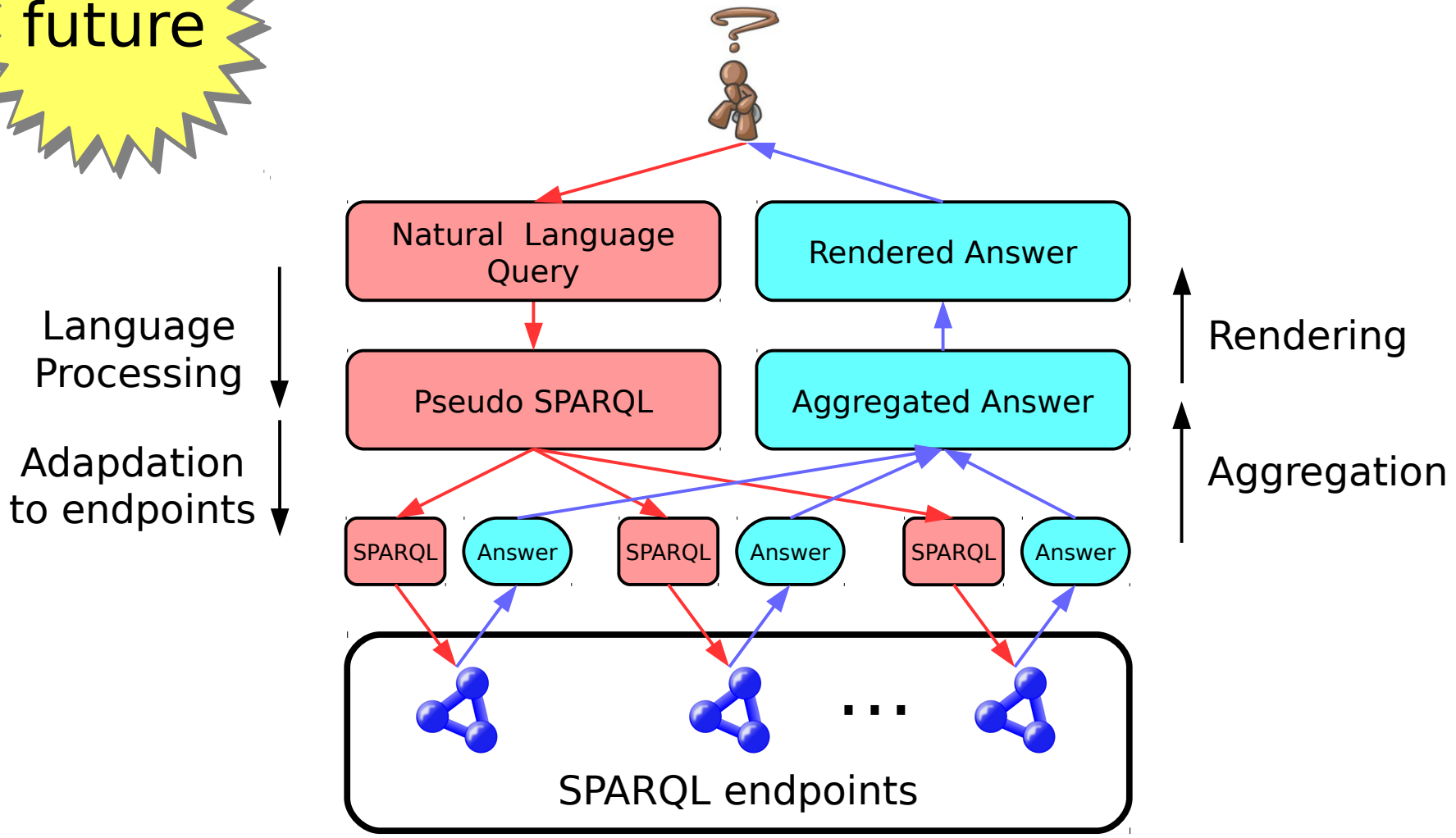
NLQA (QA on LOD)



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Federated QA on LOD

future

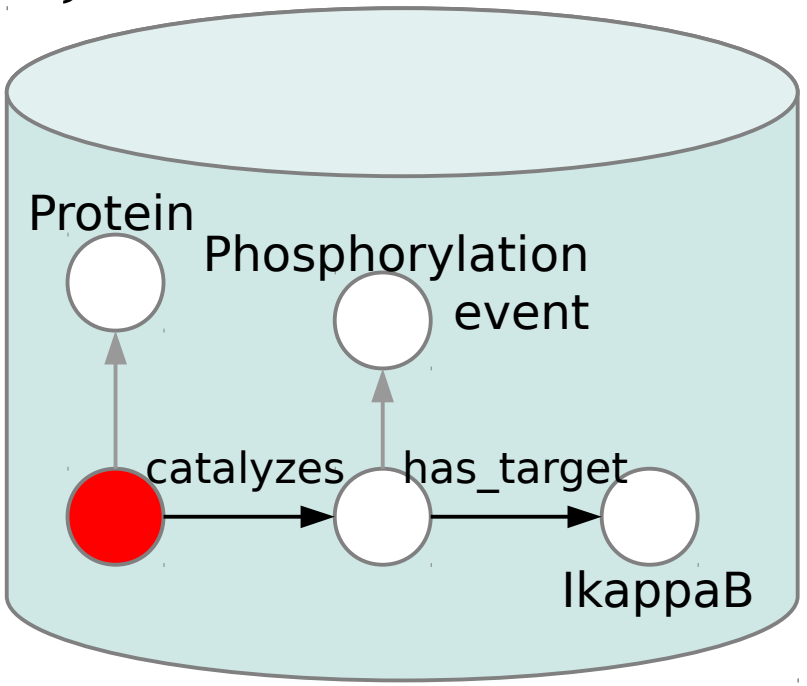


Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Challenges

- Discrepancy
  - ✔ Model representation (in NL)
  - ✔ Data representation (in EP)
  - ✔ Lexical discrepancy
  - ✔ Structural discrepancy

which proteins phosphorylate IkB?



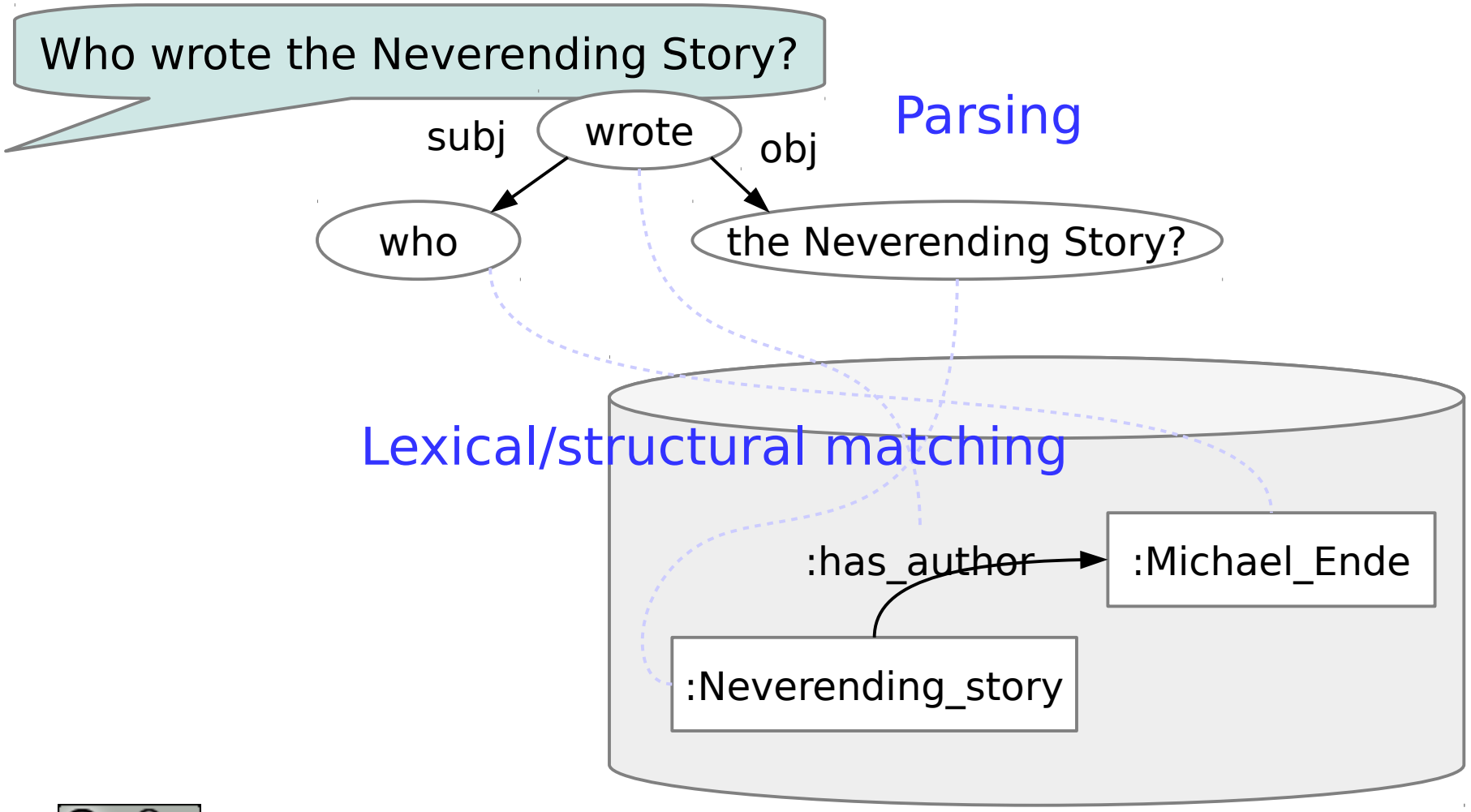
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Typical approach

- Parsing
- Lexical Matching
- Structural Matching

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Typical approach



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

LODQA

- Open source project
- Highly portable to any SPARQL endpoint
  - ✔ Assumption: SPARQL endpoints in public are beyond anybody's control.

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

LODQA

- Current state
  - ✔ Project under progress
    - ➔ Focus on addressing **structural discrepancy** (○)
    - ➔ Lexical discrepancy (△)
    - ➔ Templating (×)
    - ➔ Relation matching is not yet implemented.

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

LODQA

- Current state
  - ✔ Project under progress
    - ➔ Incomplete system, but
    - ➔ useful already to some extent.
  - ✔ *“not being perfect does not mean it's useless.”*
  - ✔ *“will keep it useful during development.”*

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

LODQA

- Three step approach
  1. Graphicator (parsing)
     - ➔ Turns a natural language query into a ***pseudo graph pattern (PGP)***
  2. Lexical mapping (dictionary lookup)
     - ➔ To anchor the PGP on the target graph
     - ➔ ***anchored PGP***
  3. GraphFinder
     - ➔ Search the KB graph for the anchored PGP.

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

LODQA

Querying linked open data (LOD) using natural language. Sounds great? Let's realize it!

Front
Motivation
Participants
References
github

LODQA (Linked Open Data Question Answering) is an open source project aiming at developing a system to generate SPARQL queries from natural language queries.

News

- (02/17-21/2014) OKBQA 2014 Hackathon is held in Jeju.
- (01/10/2014) Relation detection is improved to find the right pairs of terms in relation.
- (12/20/2013) LODQA is launched as an open source project.

Prototype Demo (targeting OMIM)

what side effects are associated with streptomycin? submit

Example

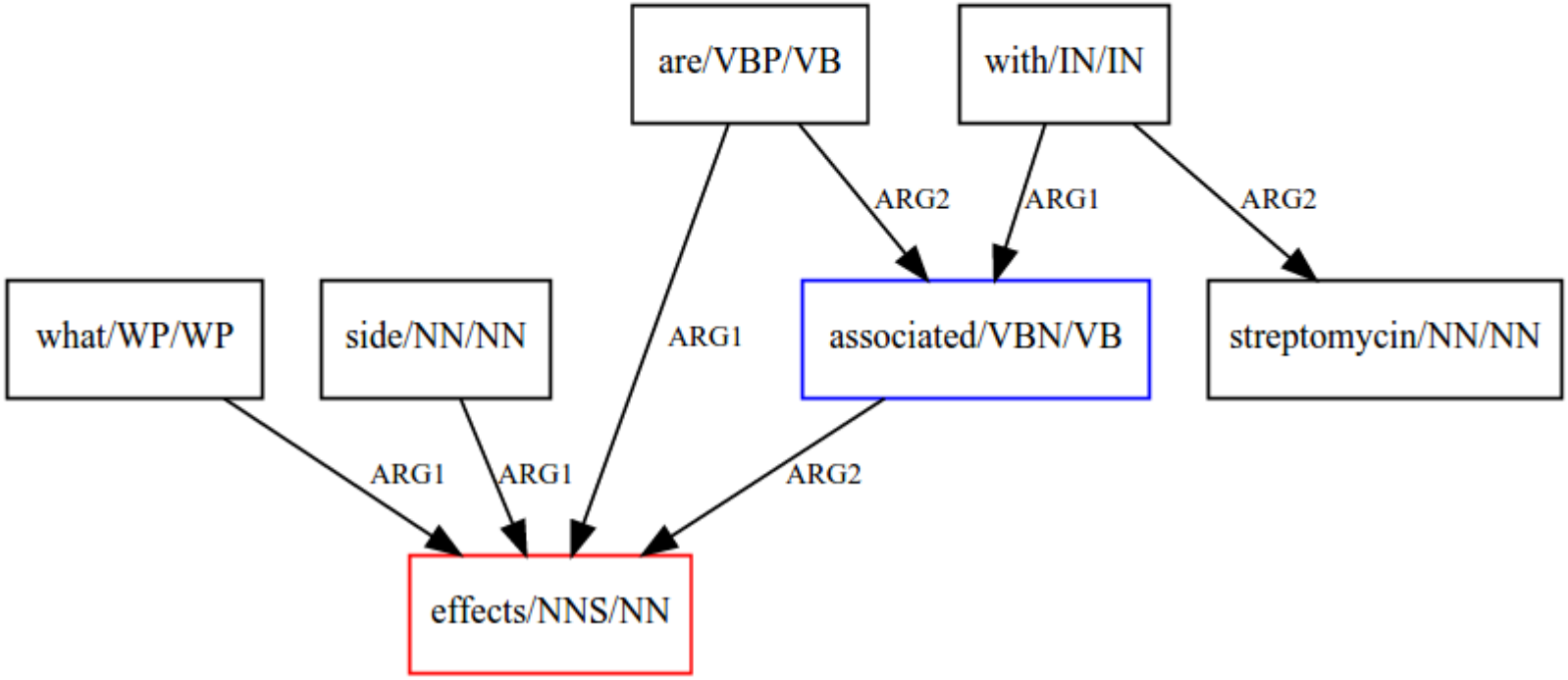
- what genes are associated with kabuki syndrome?
- what sign is associated with kabuki syndrome?
- what cellular dysfunction is associated with kabuki syndrome?
- what pathologic function is associated with kabuki syndrome?
- what neoplastic process is associated with kabuki syndrome?
- what anatomical abnormality is associated with kabuki syndrome?
- what genes are associated with alzheimer disease?

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Query

what side effects are associated with streptomycin?

Semantic analysis (predicate-argument relation graph)



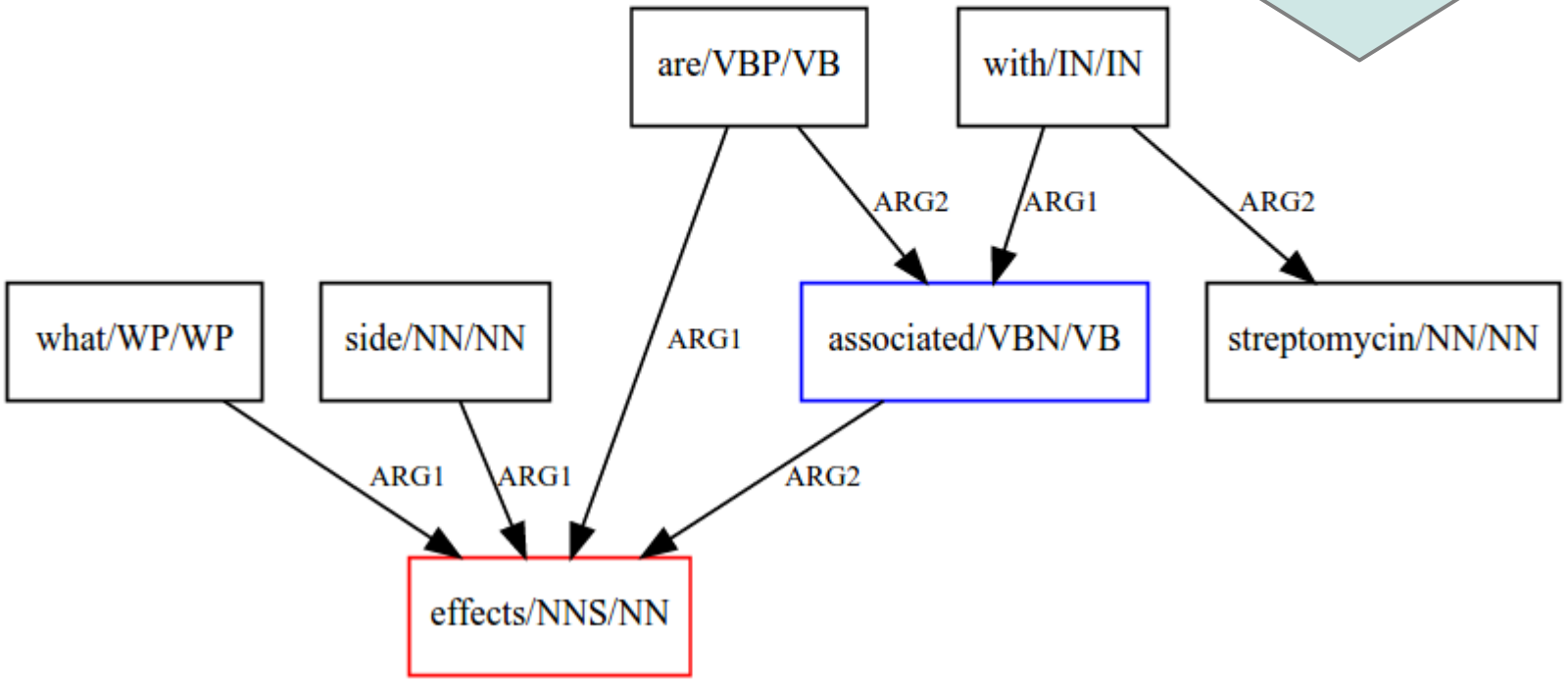
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Query

what side effects are associated with streptomycin?

Enju HPSG parser

Semantic analysis (predicate-argument relation graph)

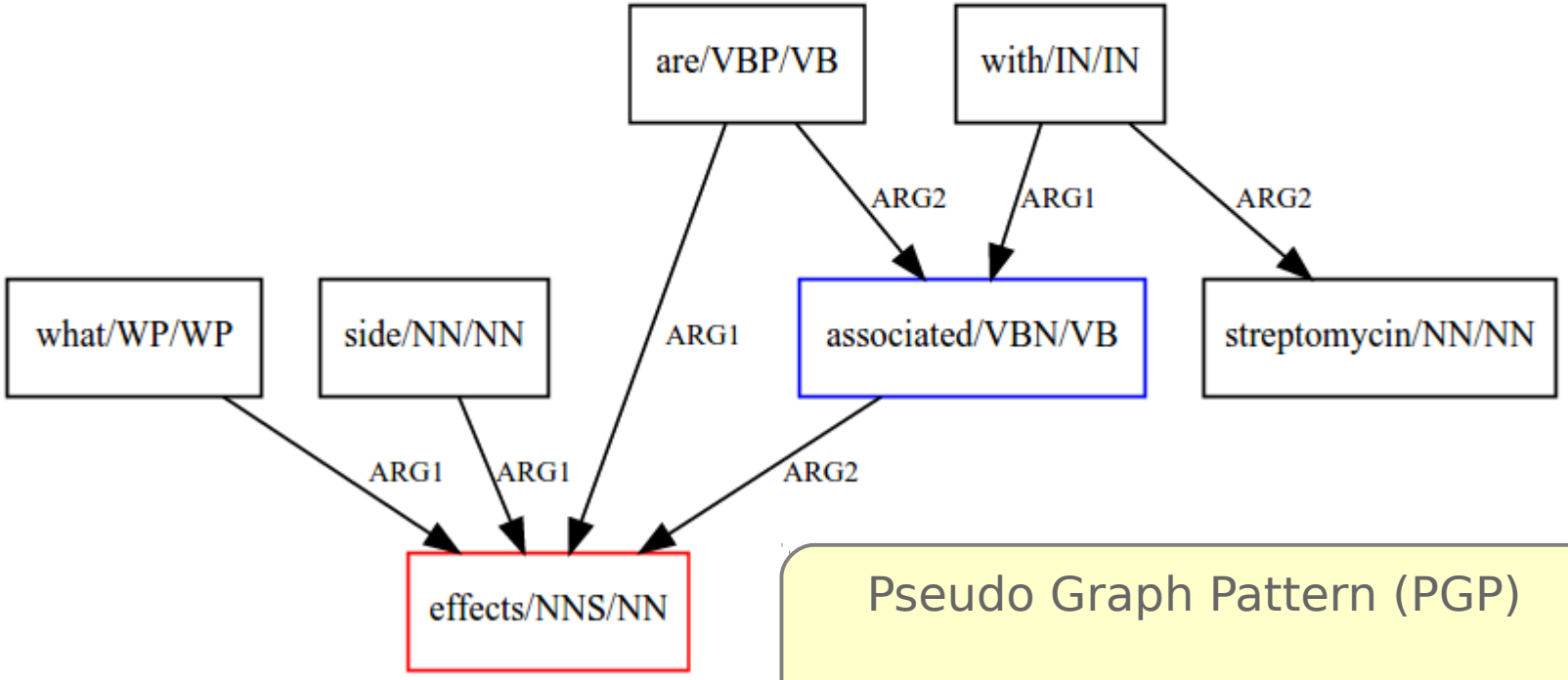


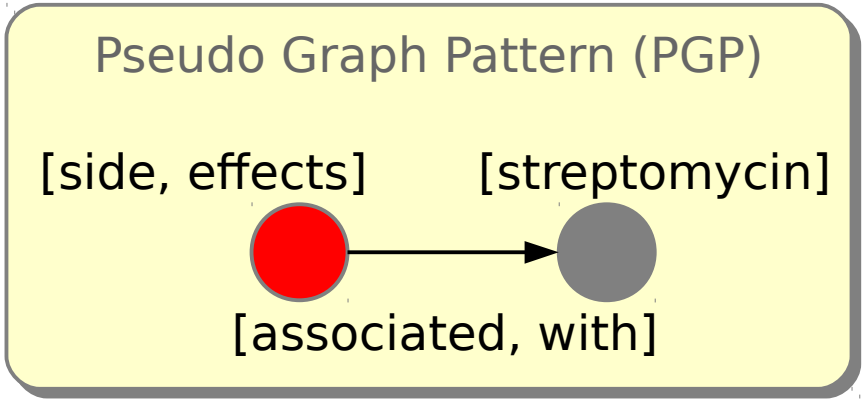
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Query

what side effects are associated with streptomycin?

Semantic analysis (predicate-argument relation graph)



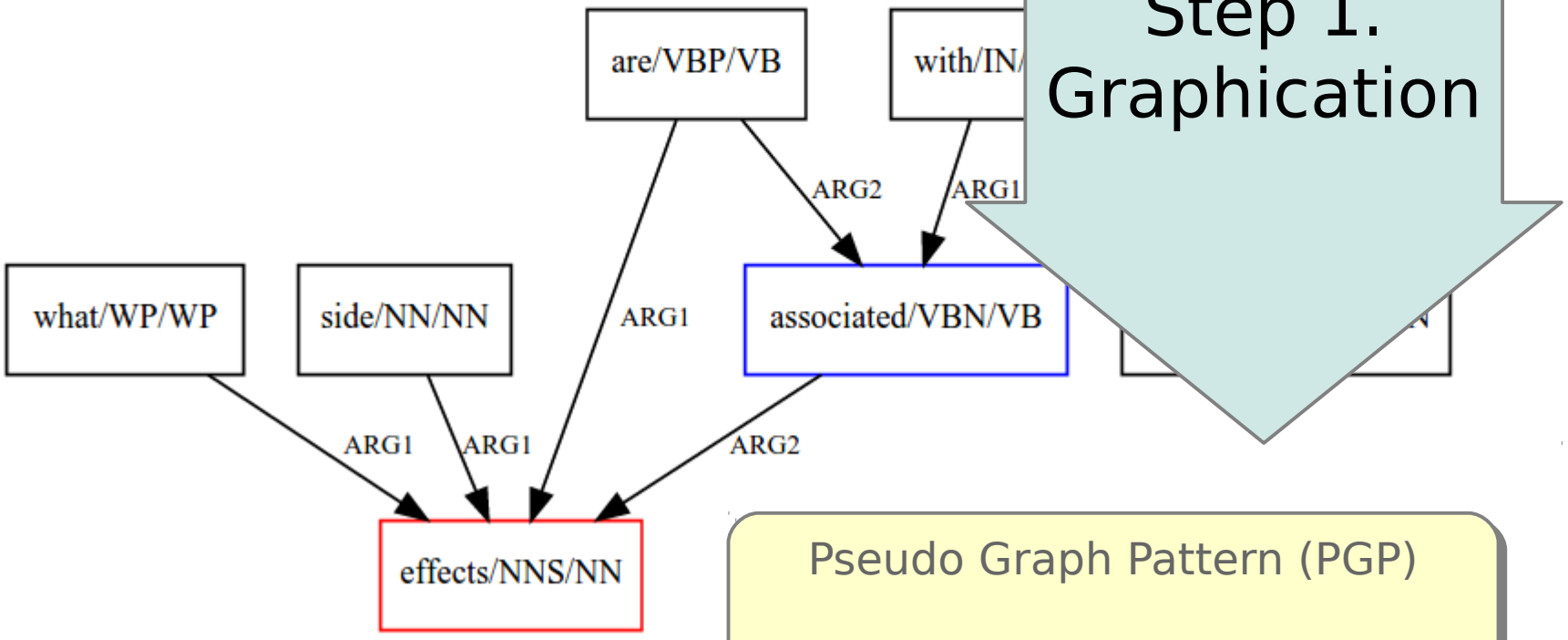


Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Query

what side effects are associated with streptomycin?

Semantic analysis (predicate-argument relation graph)



Step 1.
Graphication

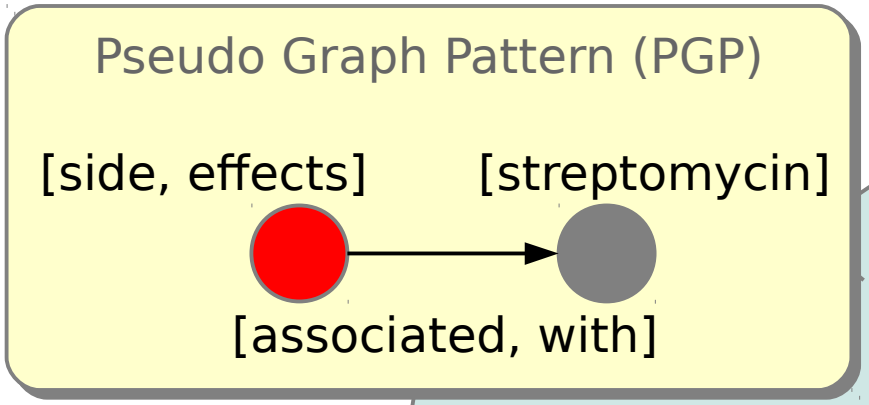
Pseudo Graph Pattern (PGP)

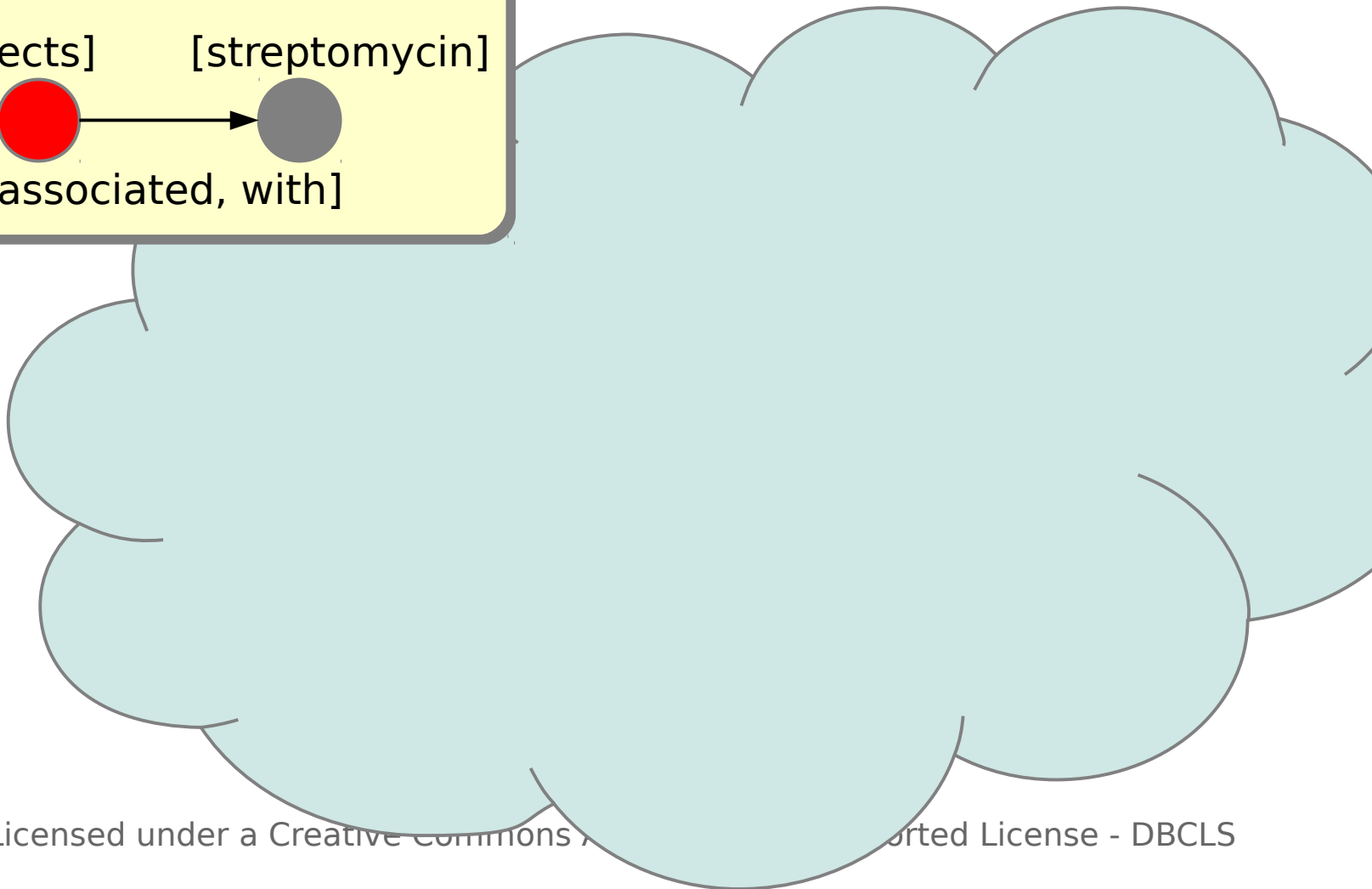
[side, effects] [streptomycin]

[associated, with]

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Graph Pattern matching





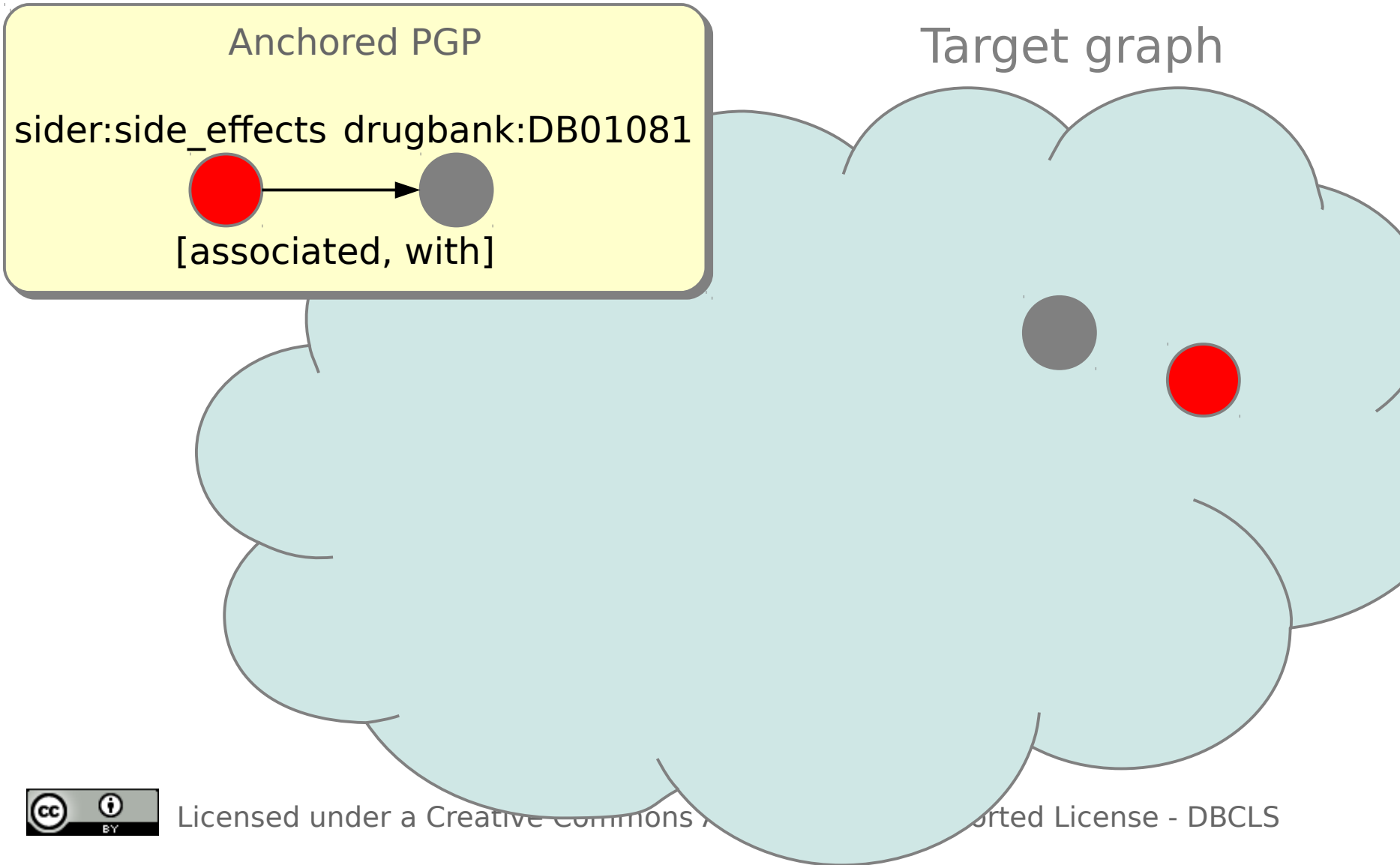
Licensed under a Creative Commons A orted License - DBCLS

Step 2. Lexical Mapping

- [side, effect]
  - ✔ sider:side_effects
  - ✔ sider:sideEffectName
- [streptomycin]
  - ✔ drugbank:DB01082
  - ✔ drugbank:DB00428
  - ✔ Sider:5297
  - ✔ sider:5300

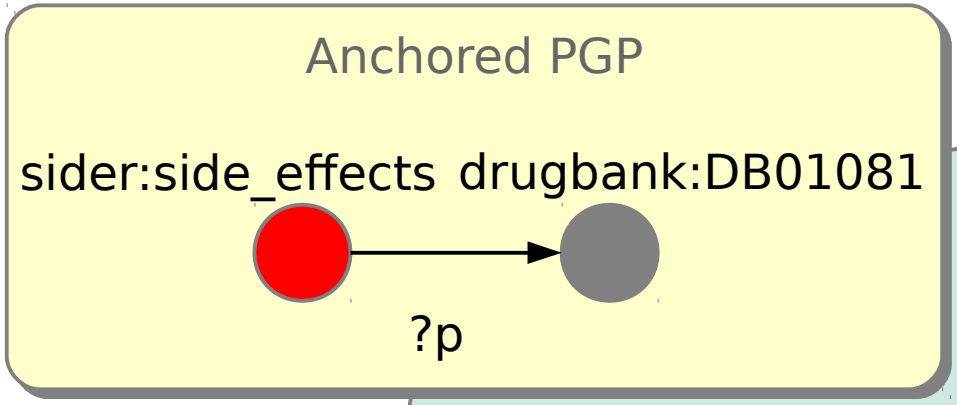
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

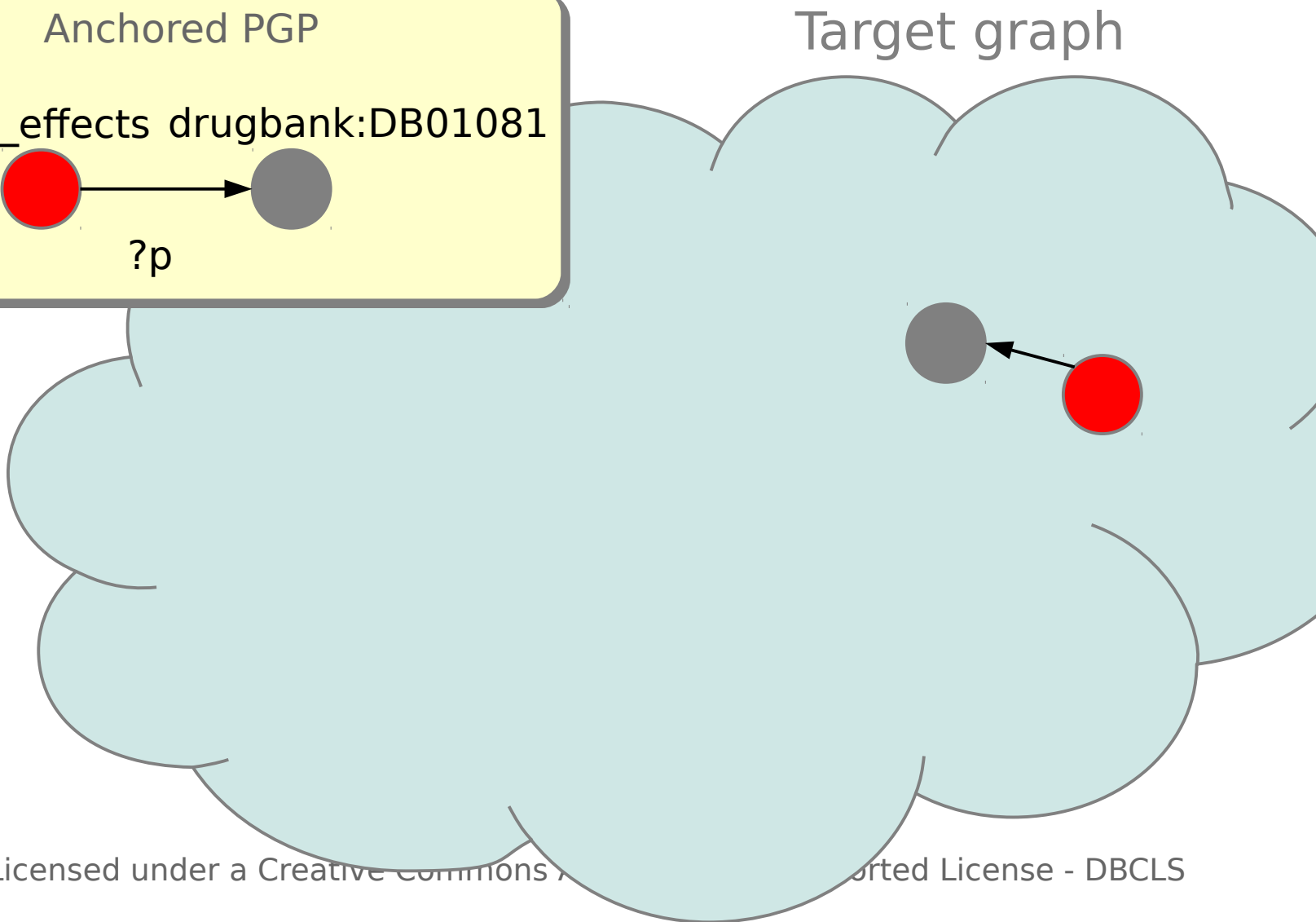
Step 3. GraphFinder



Licensed under a Creative Commons A orted License - DBCLS

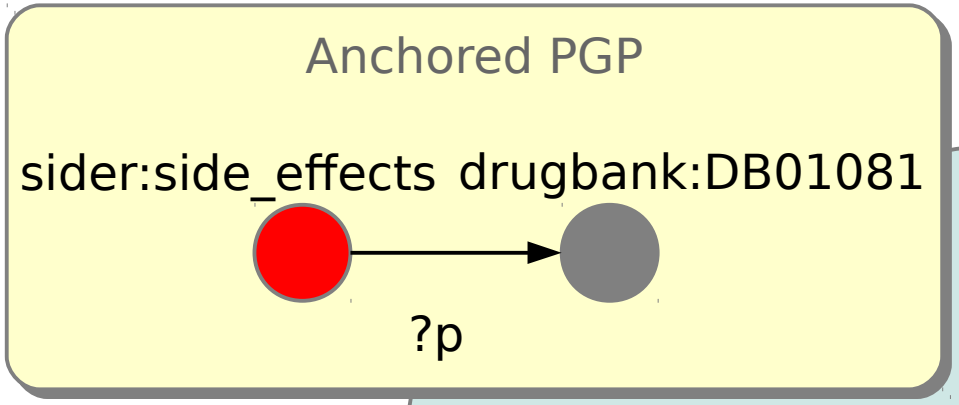
Step 3. GraphFinder

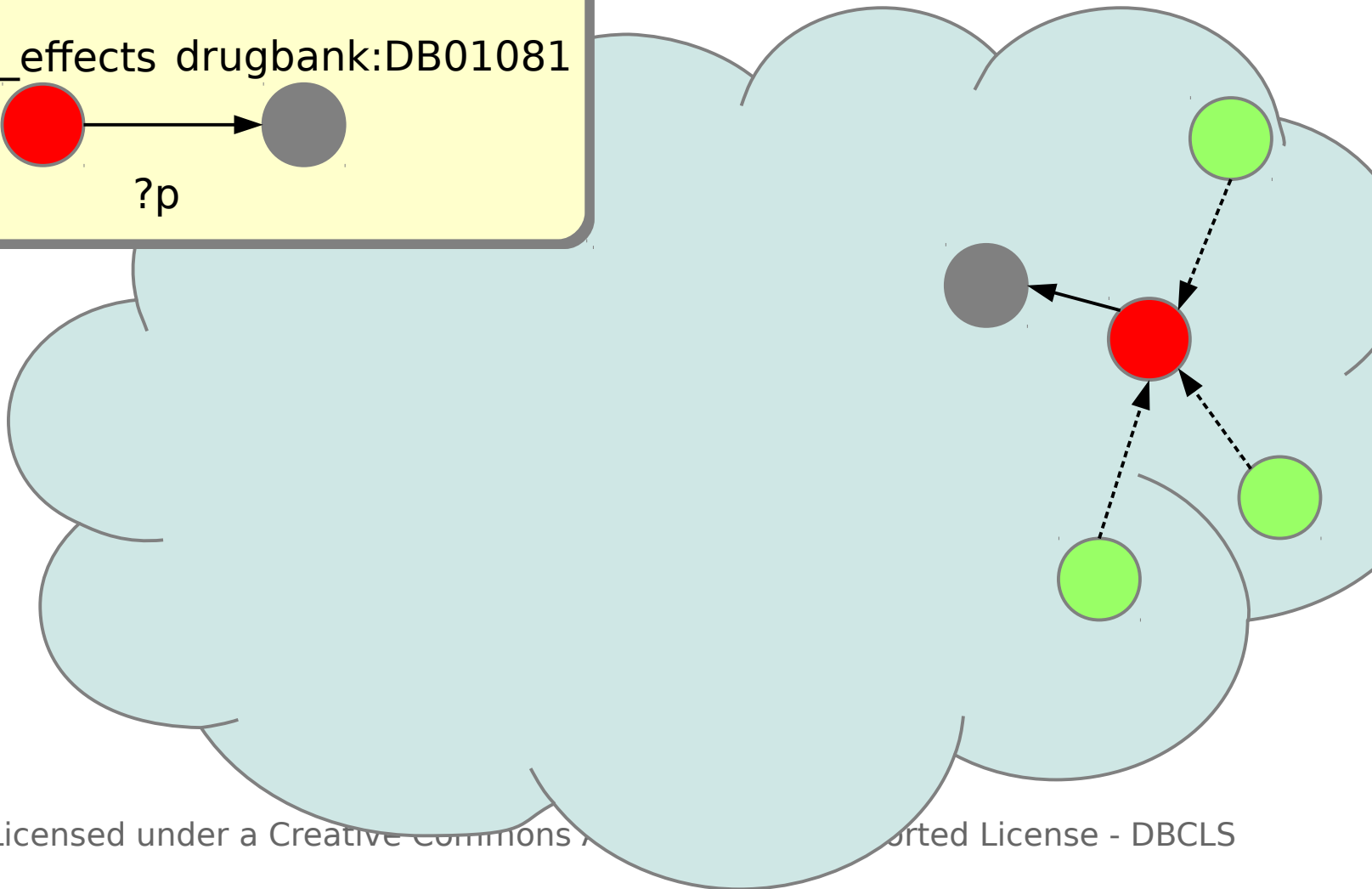




Licensed under a Creative Commons A orted License - DBCLS

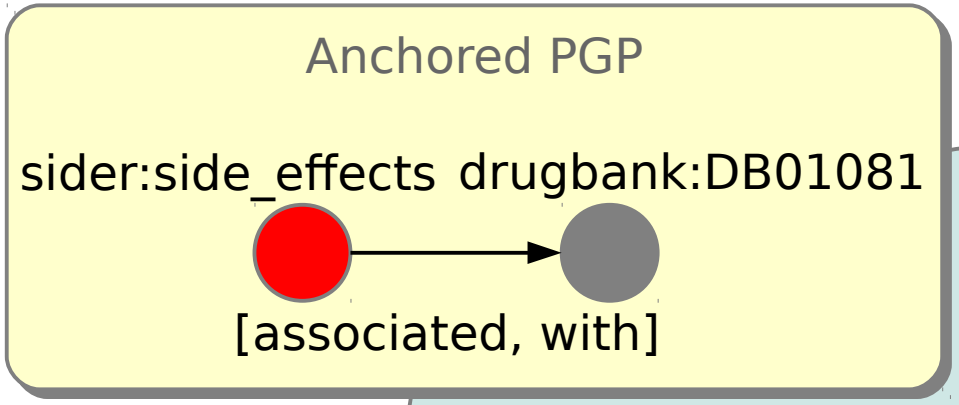
Final output: instances of the focused node

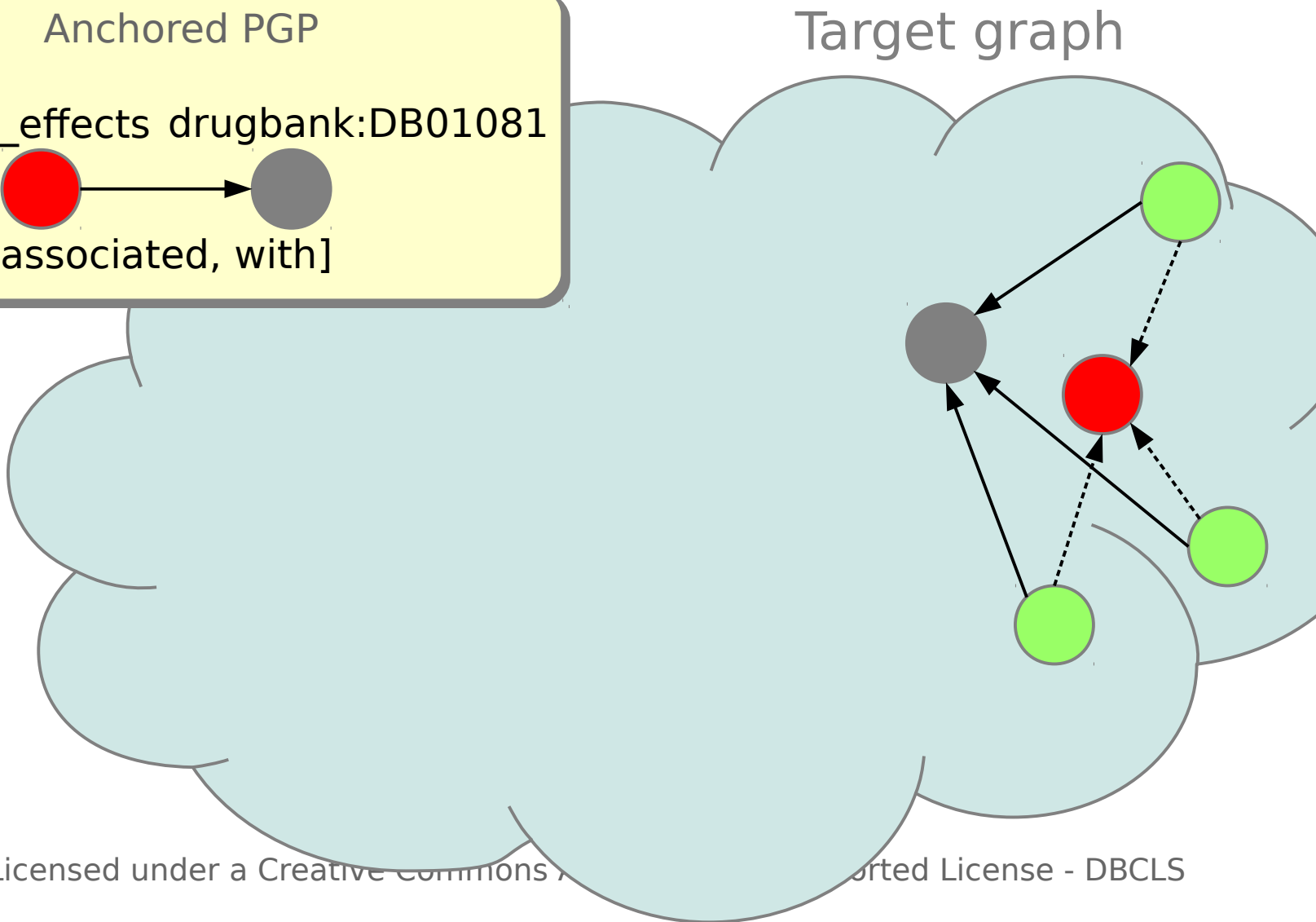




Licensed under a Creative Commons A orted License - DBCLS

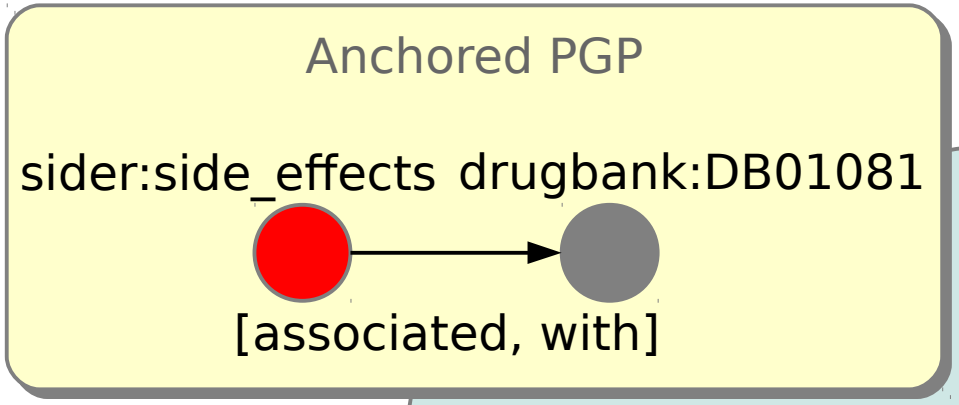
Representational variations

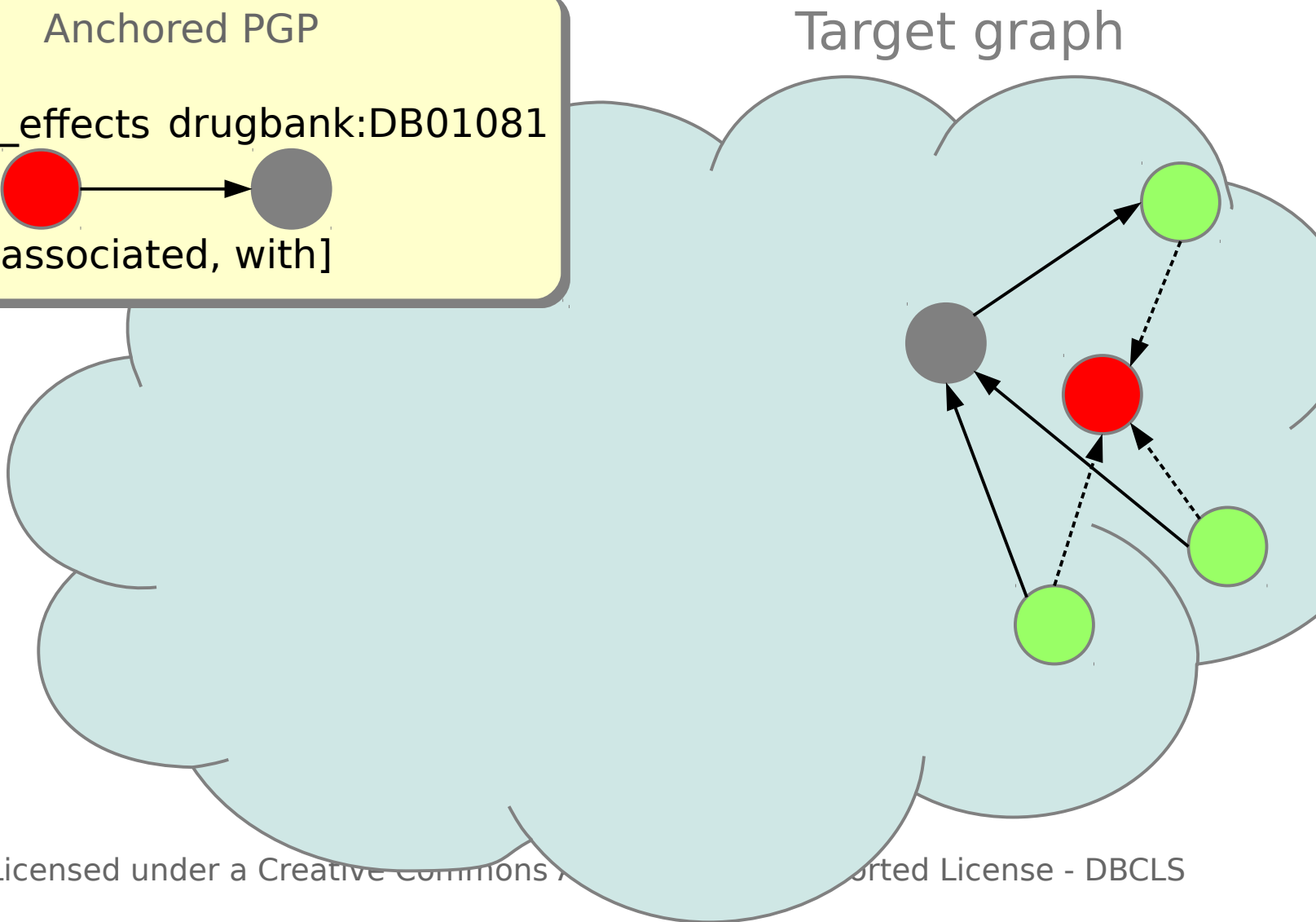




Licensed under a Creative Commons A orted License - DBCLS

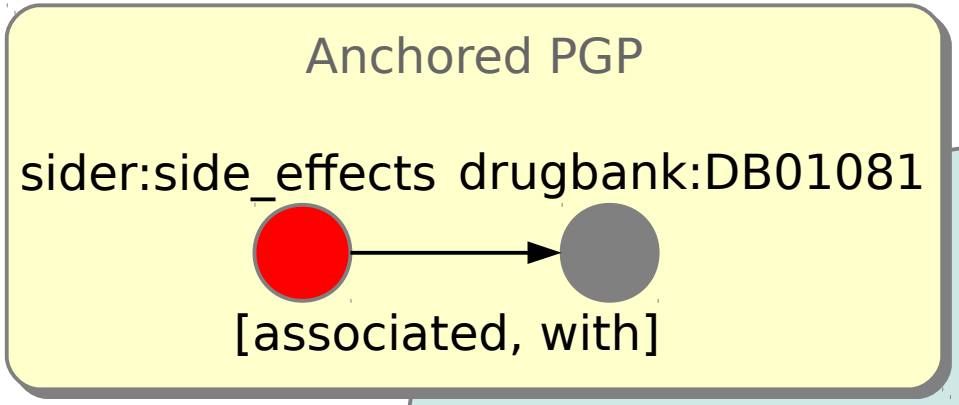
Representational variations

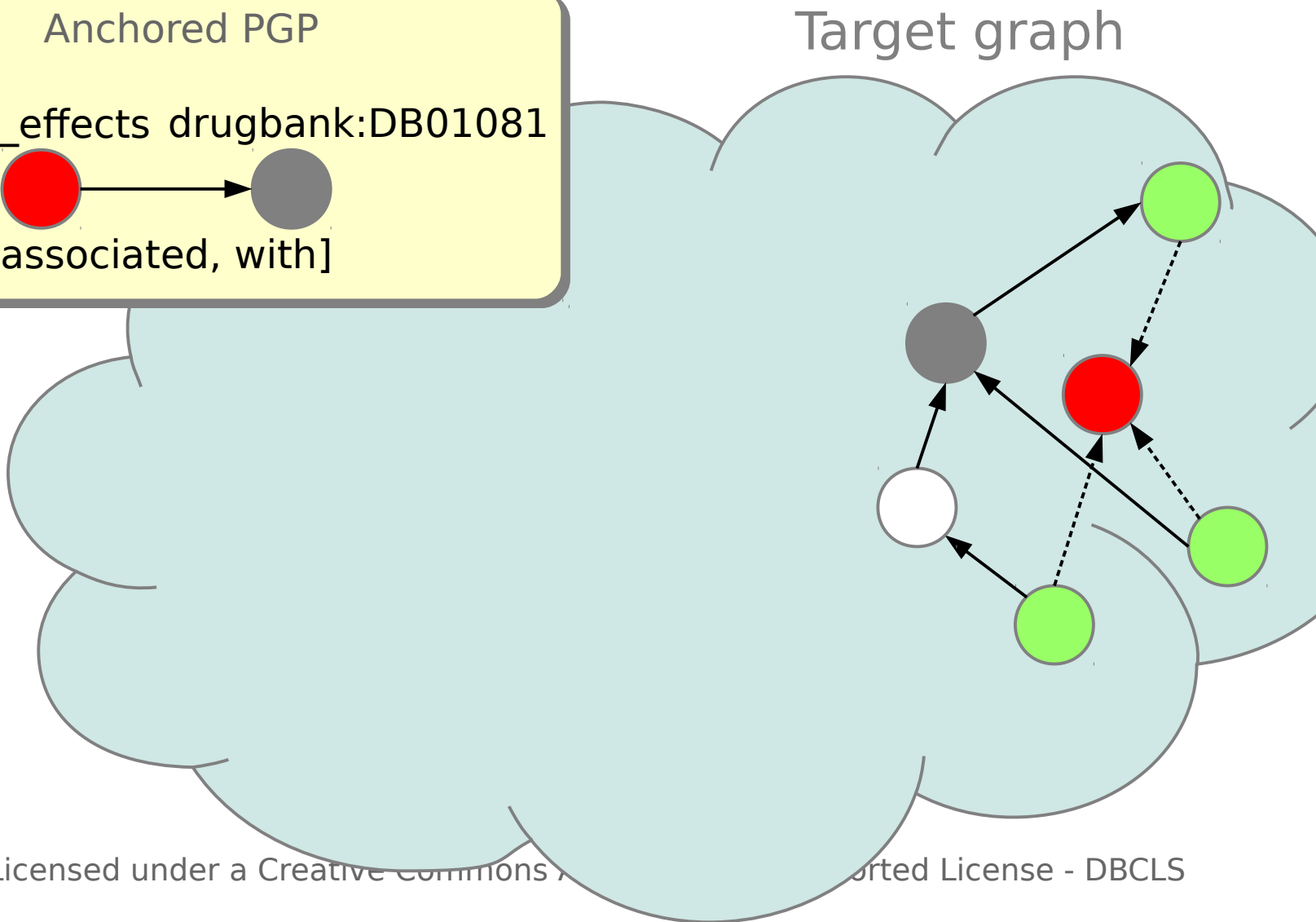




Licensed under a Creative Commons A orted License - DBCLS

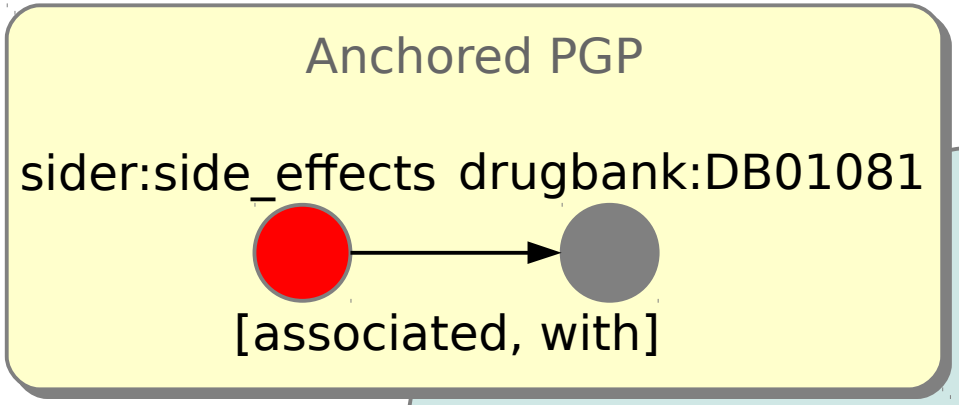
Representational variations

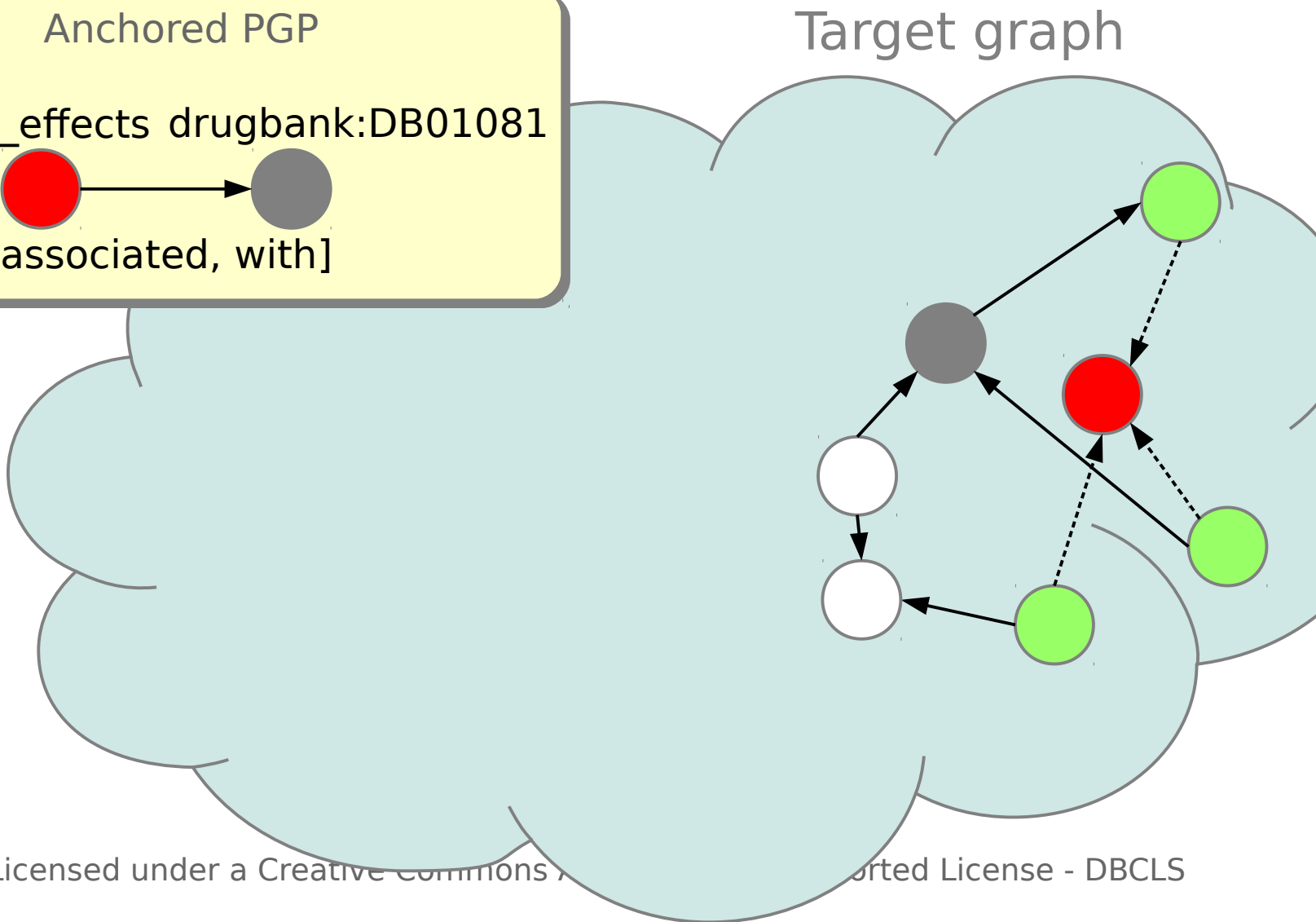




Licensed under a Creative Commons A orted License - DBCLS

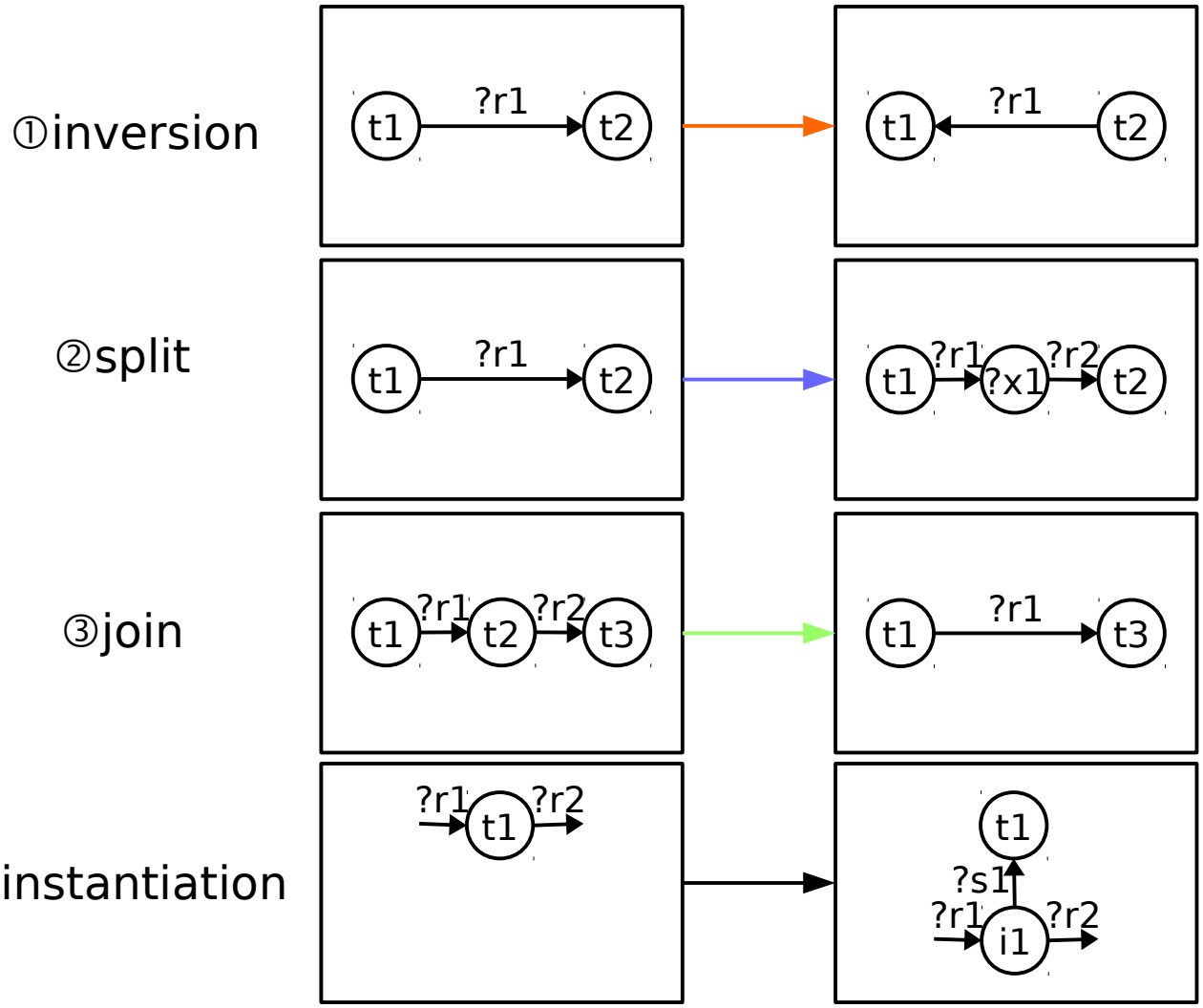
Representational variations





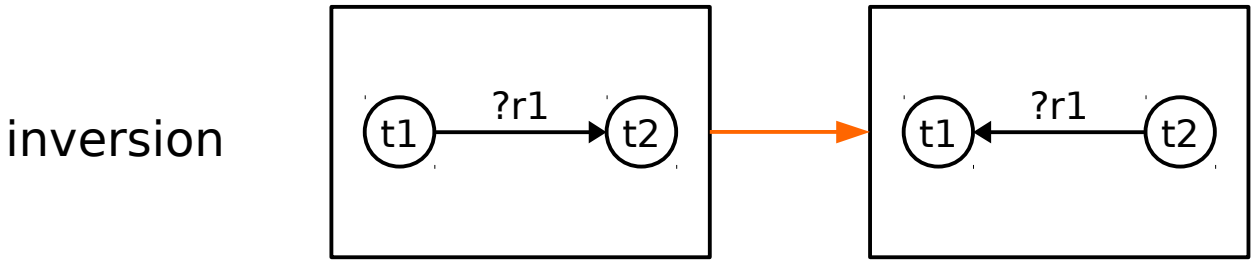
Licensed under a Creative Commons A orted License - DBCLS

Operations for graph variation

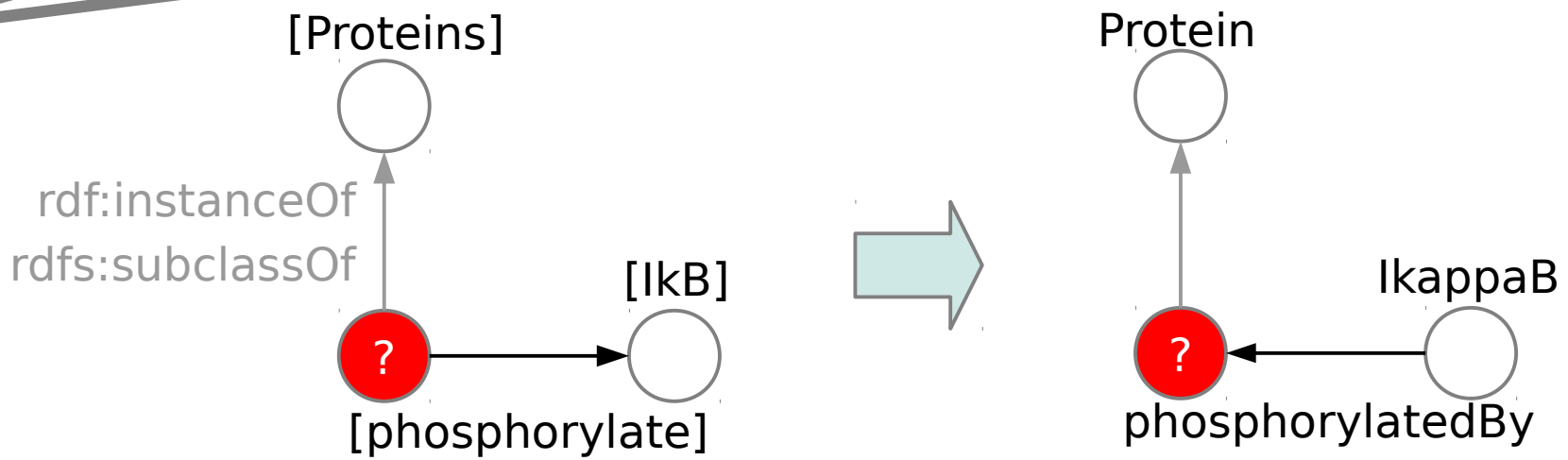


Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

①Inversion

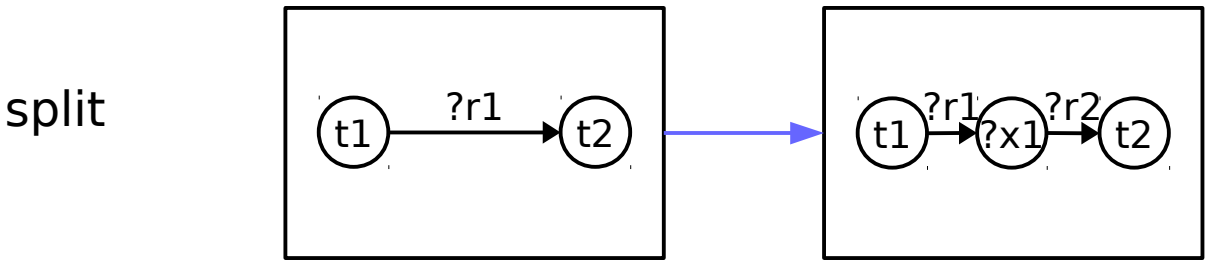


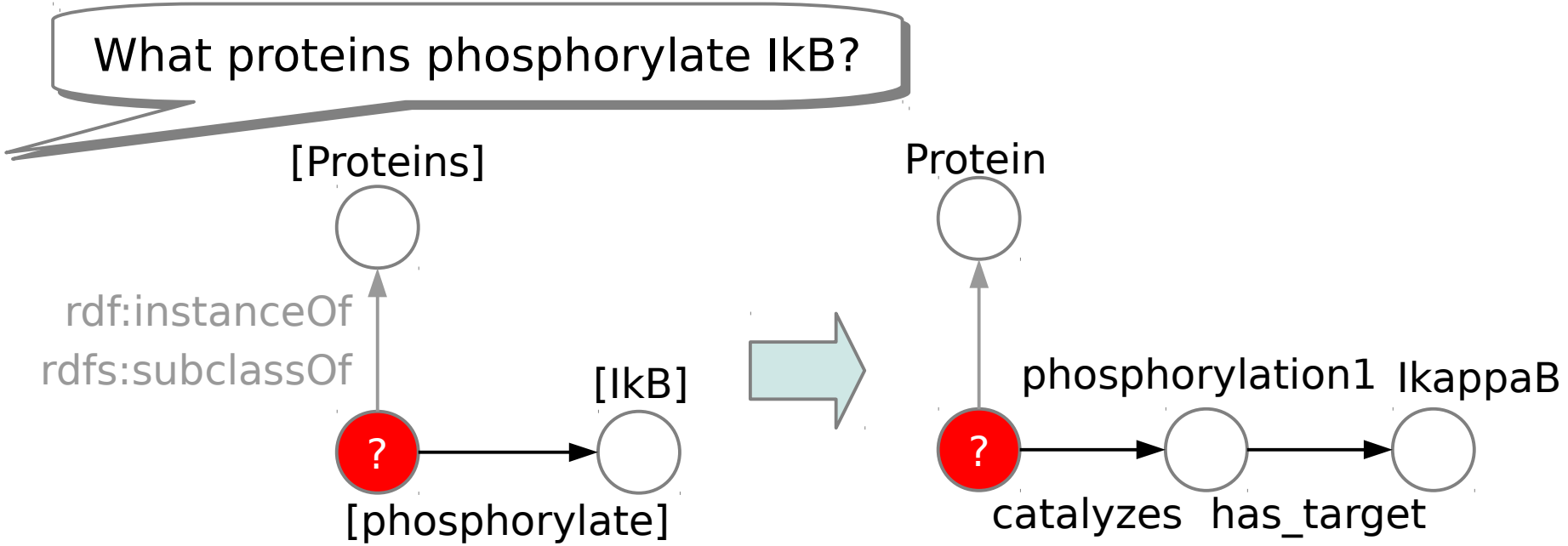
What proteins phosphorylate IkB?



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

②Split

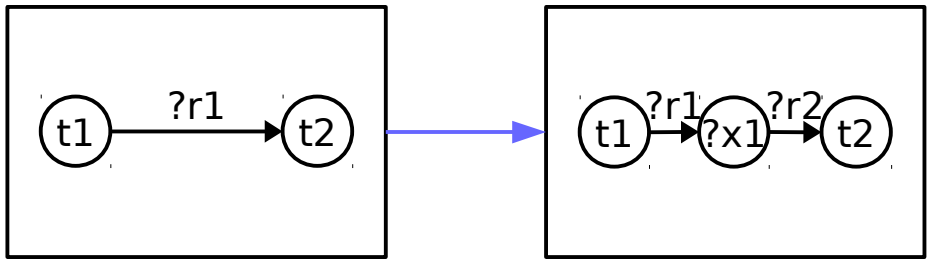




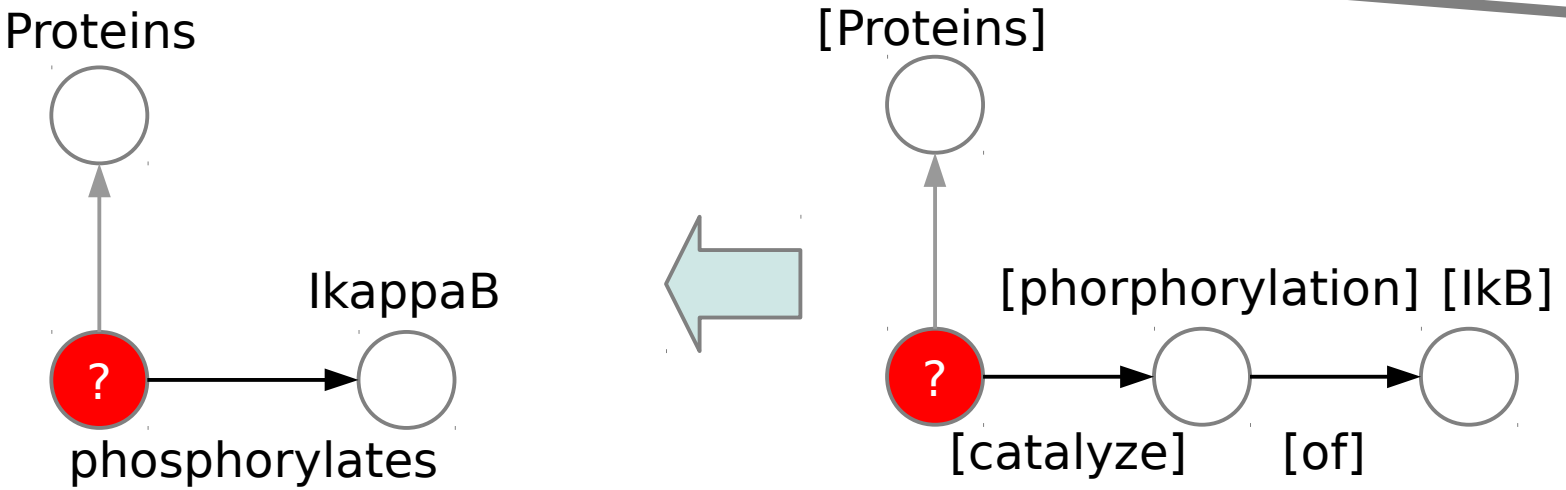
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

③Join

split



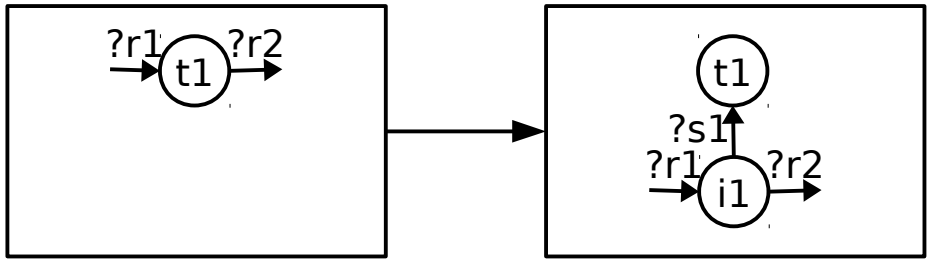
What proteins catalyze the phosphorylation of IkB?



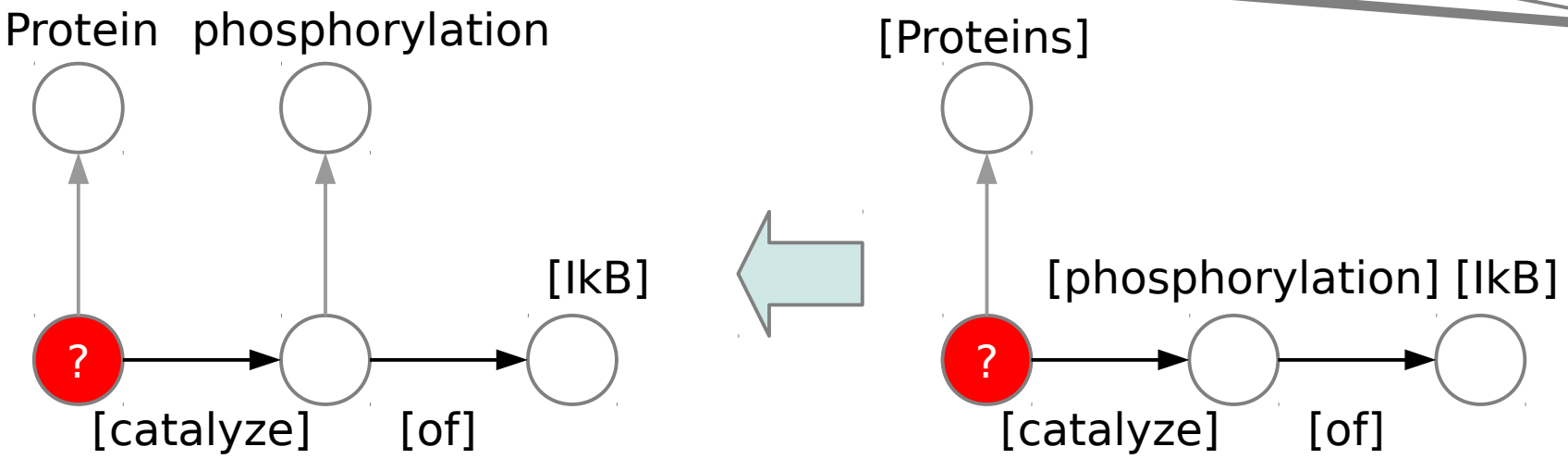
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

④Instantiation

instantiation

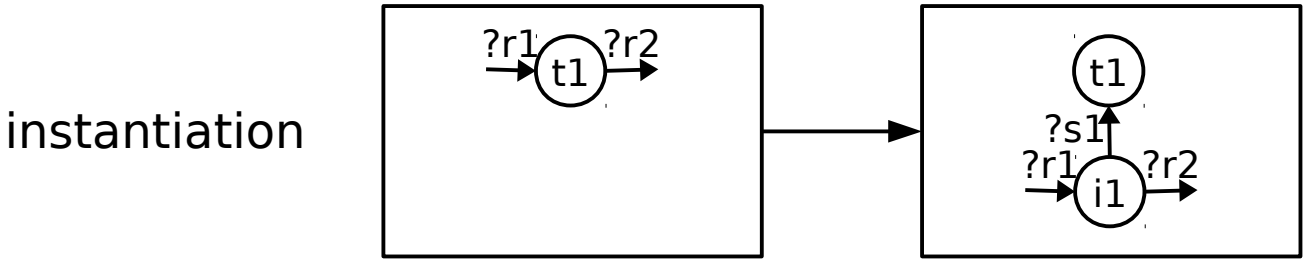


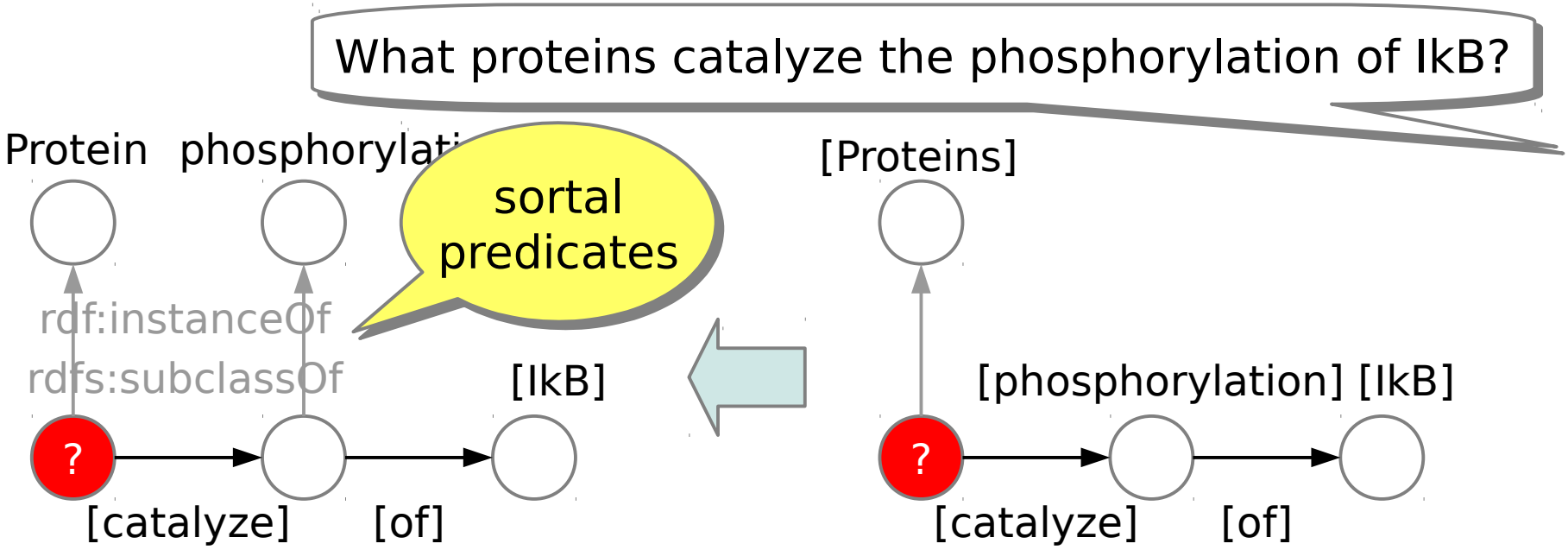
What proteins catalyze the phosphorylation of IkB?



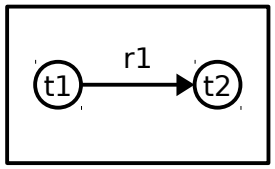
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

④Instantiation

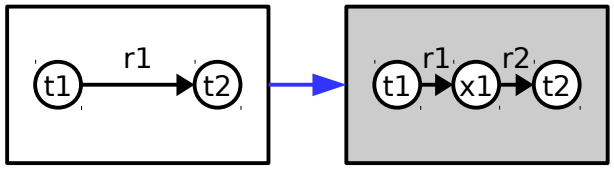




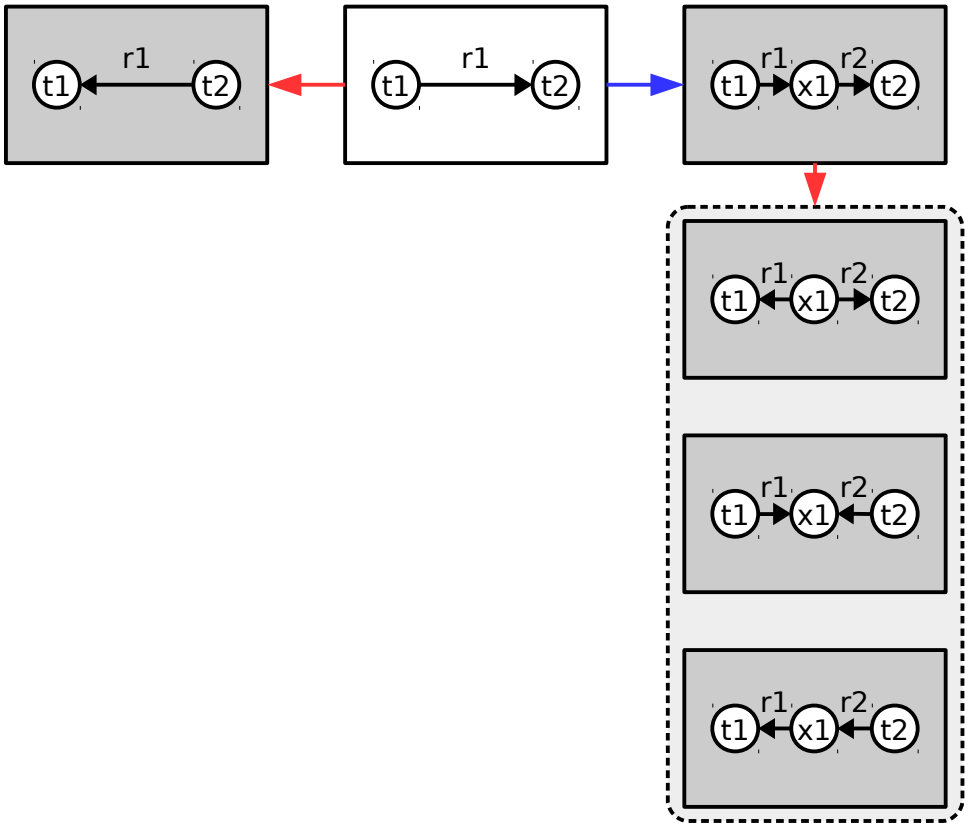
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



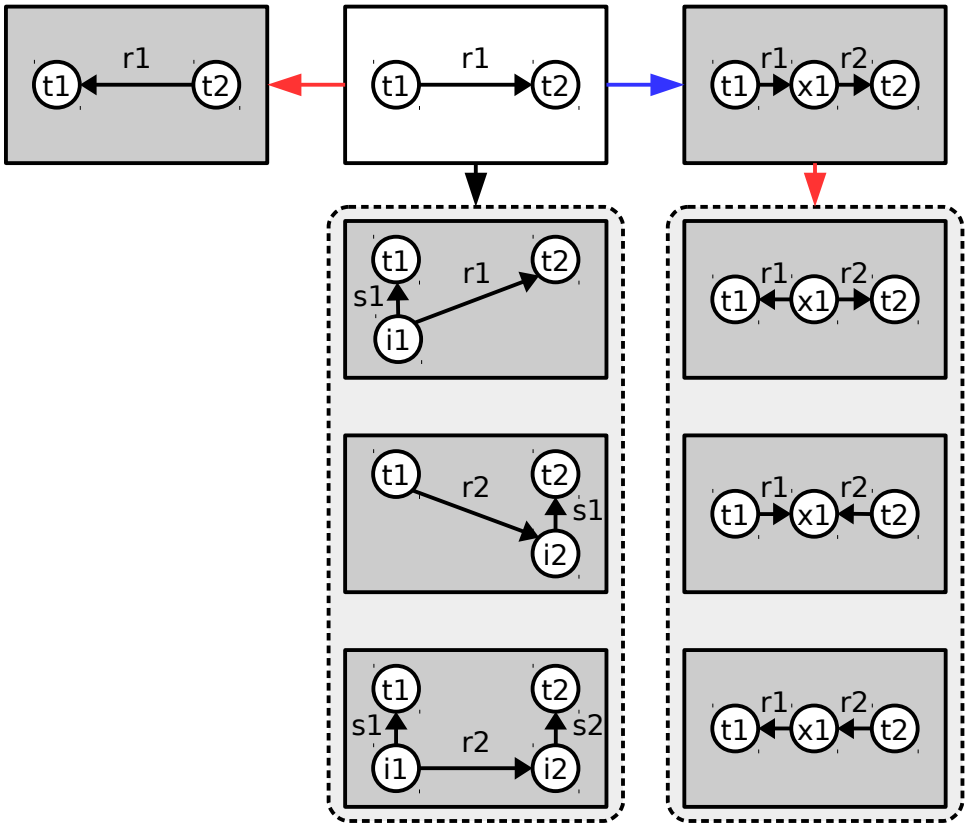
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



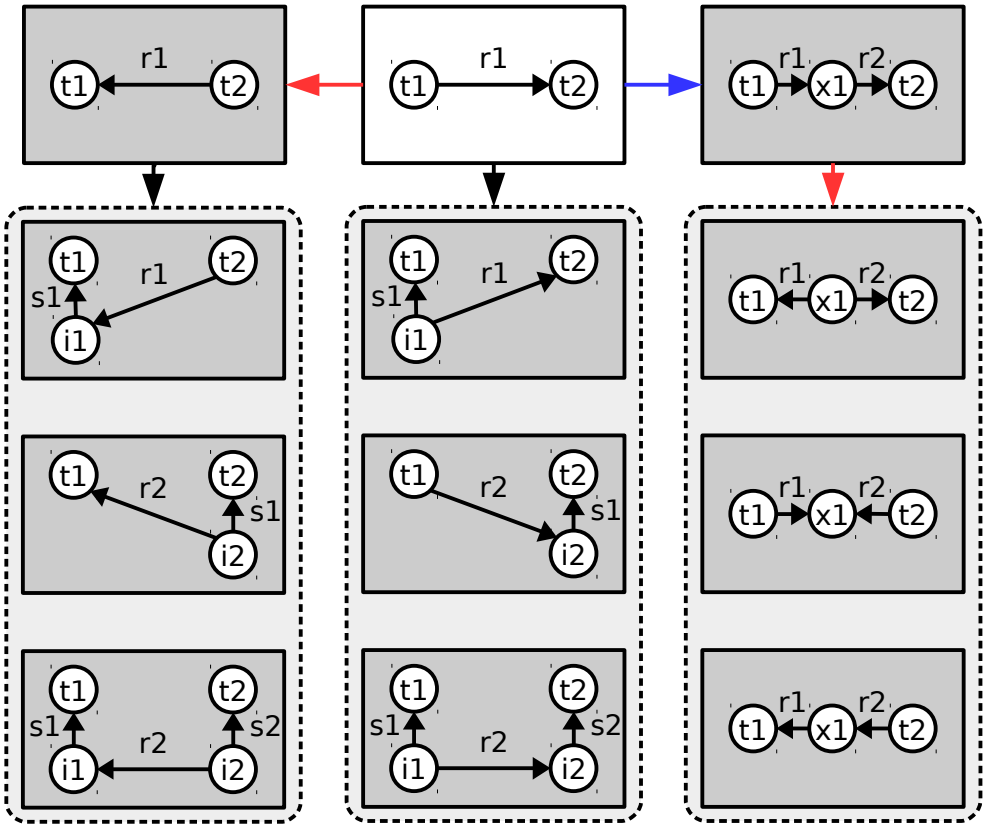
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



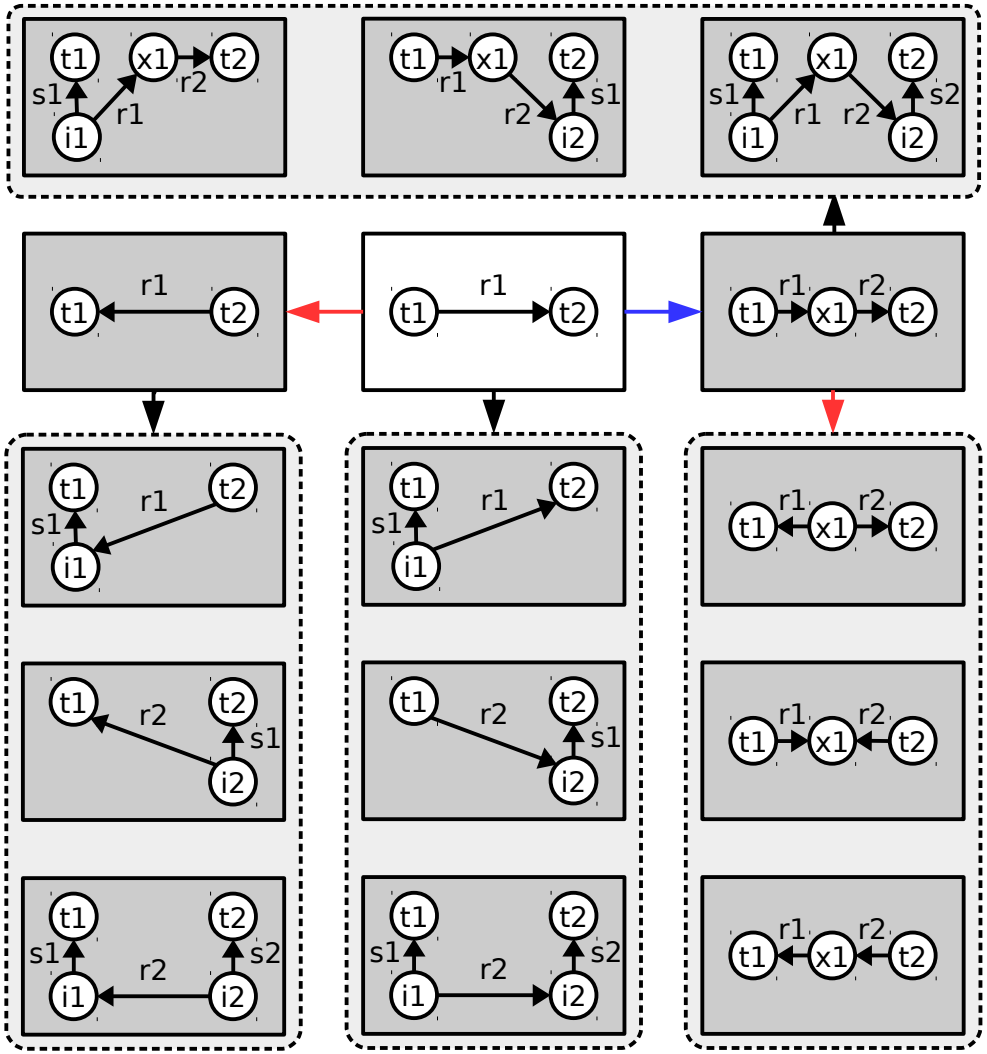
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



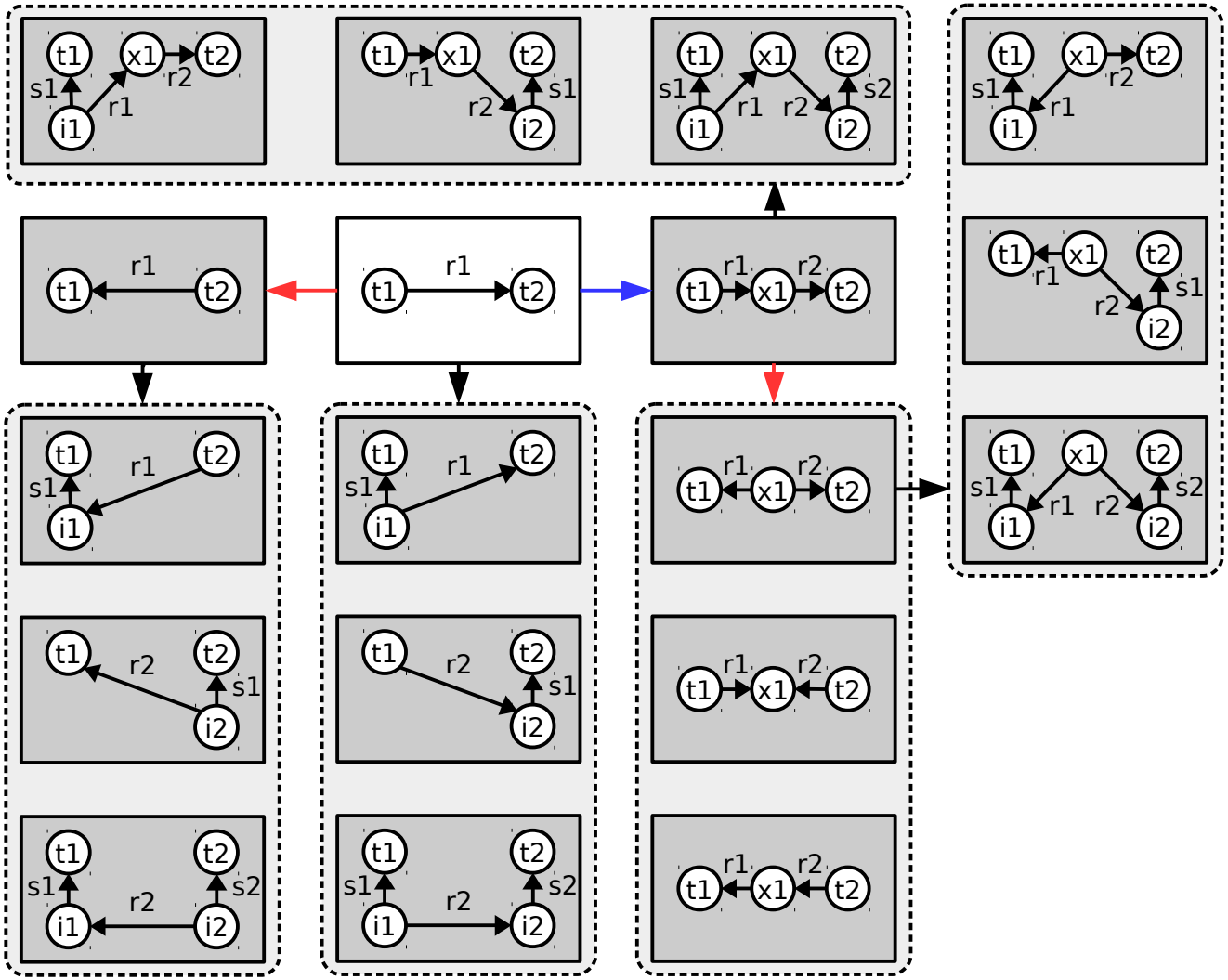
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



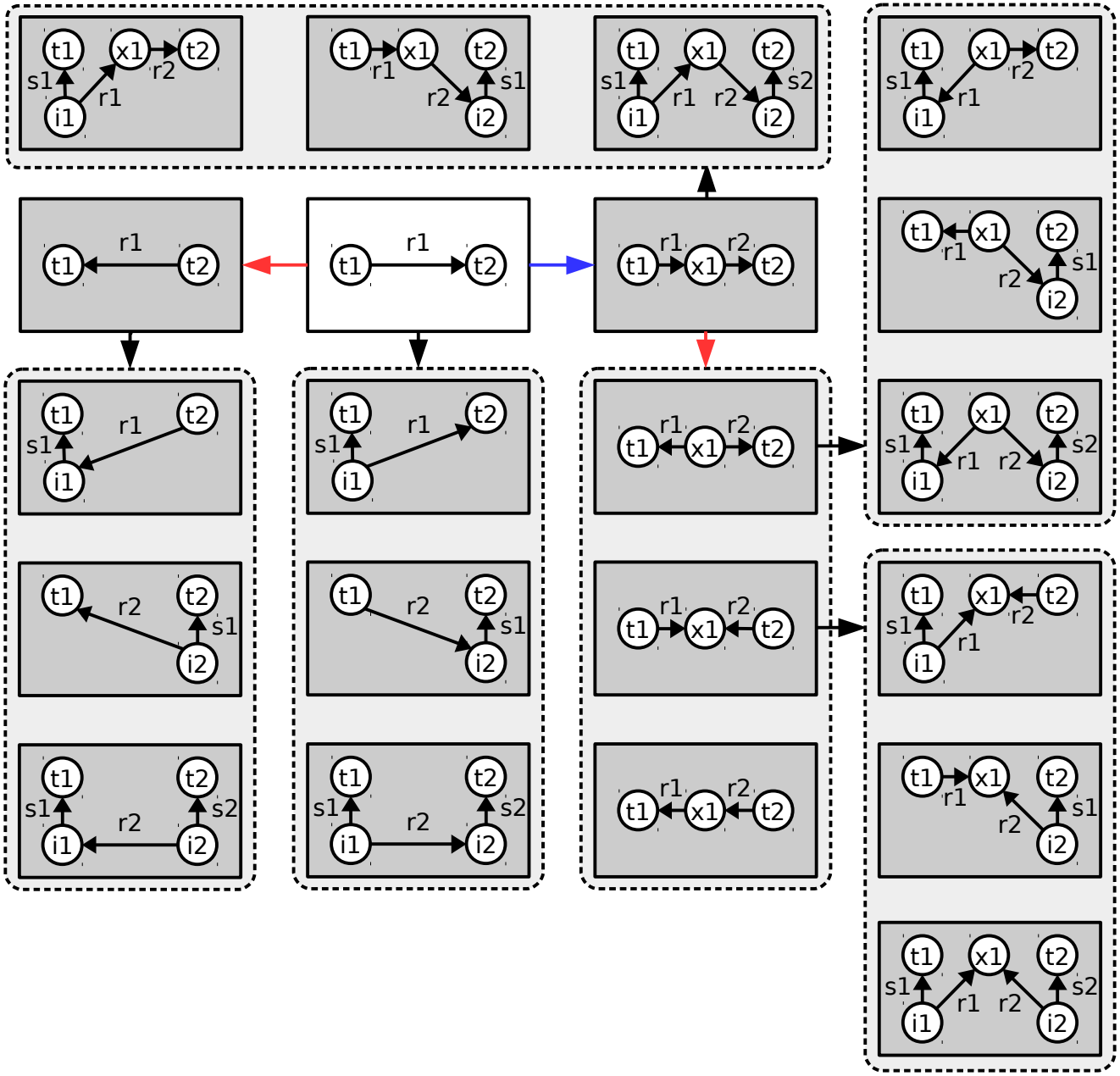
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



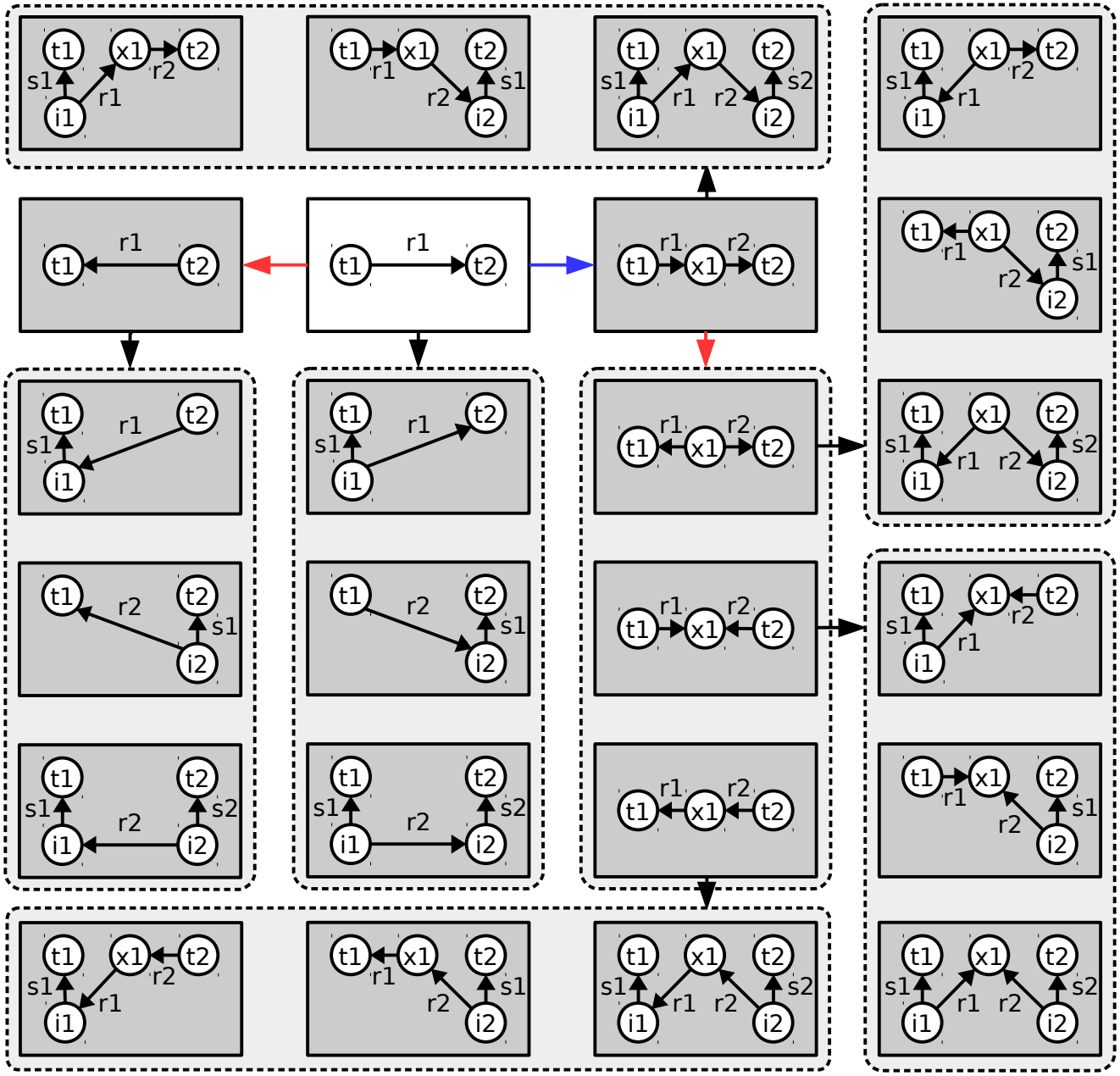
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



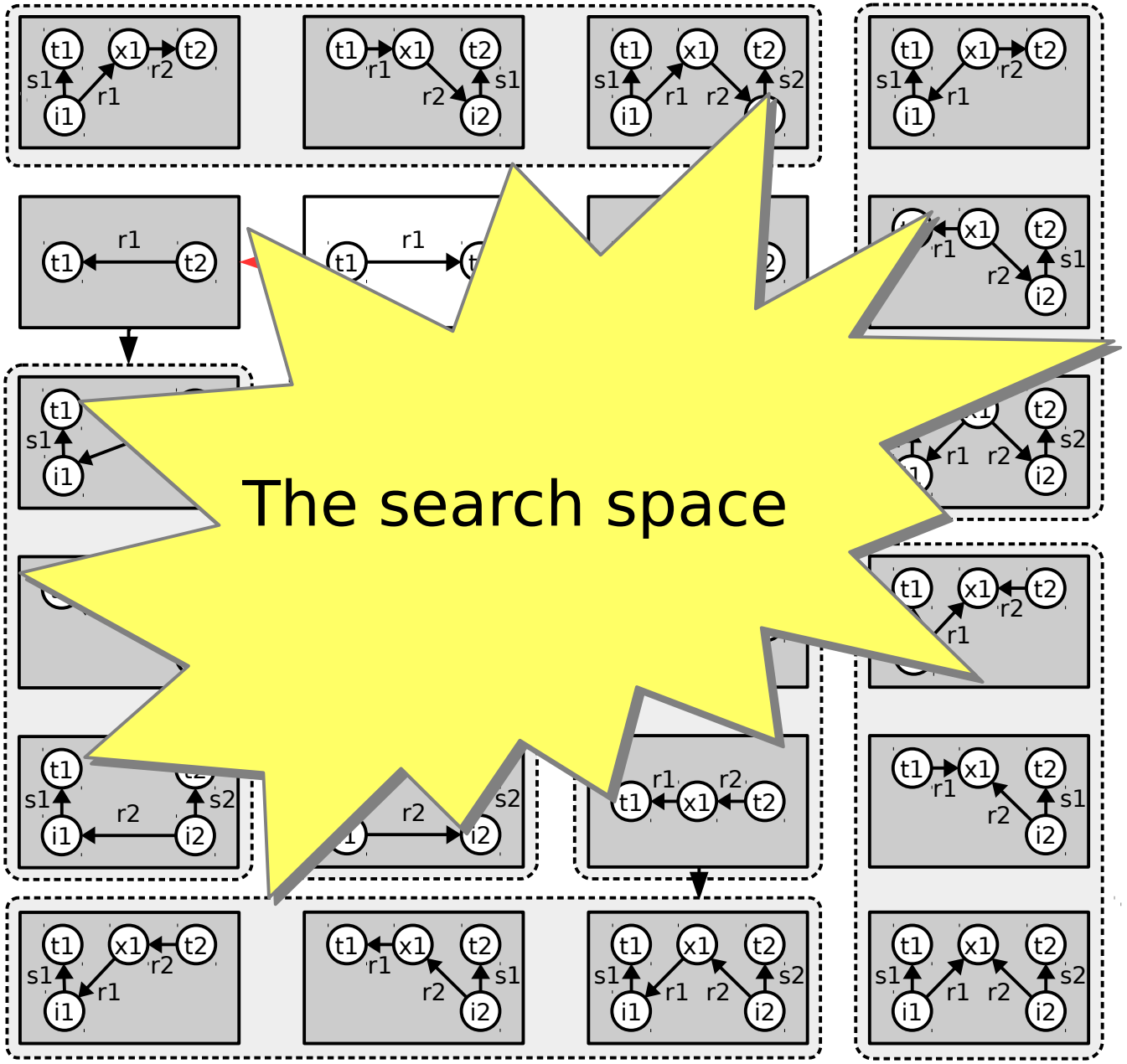
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Demo

- http://www.lodqa.org

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Comparison to RelFinder

- RelFinder
  - ✔ http://www.visualdataweb.org/relfinder.php
- GraphFinder generalizes RelFinder
  - ✔ two instances → two, three, four, ...
  - ✔ → classes or instances

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

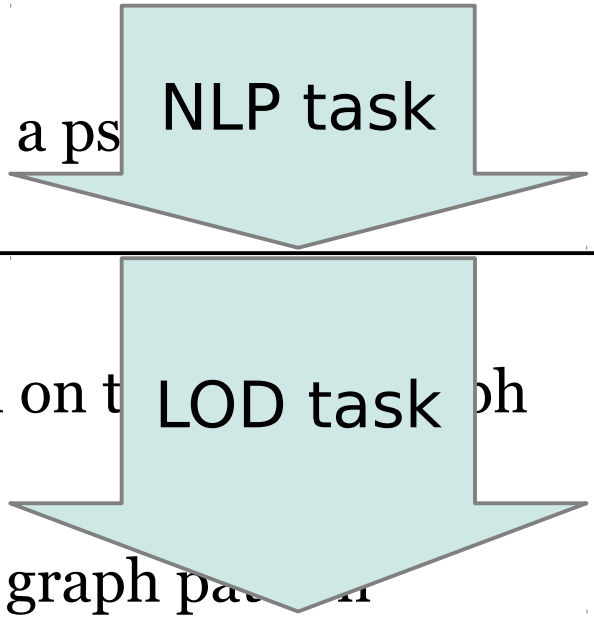
Summary

- Three step approach
  1. Graphicator
     - ➔ Turns a natural language query into a pseudo graph pattern
  2. Lexical mapping
     - ➔ To anchor the pseudo graph pattern on the target graph
  3. GraphFinder
     - ➔ Search the KB graph for the pseudo graph pattern

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Summary

- Three step approach
  1. Graphicator
     - ➔ Turns a natural language query into a ps pattern

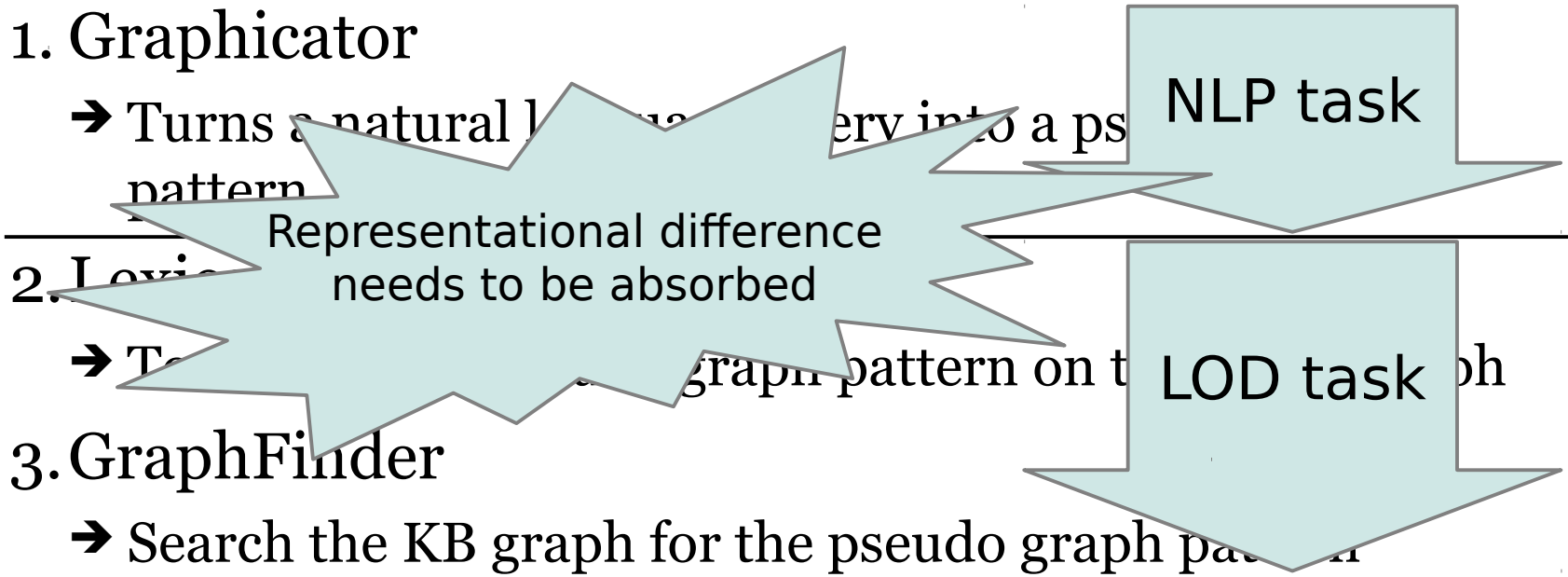


  2. Lexical mapping
     - ➔ To anchor the pseudo graph pattern on t ph
  3. GraphFinder
     - ➔ Search the KB graph for the pseudo graph pa n

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Summary

- Three step approach
  1. Graphicator
     - ➔ Turns a natural l… …ery into a ps… pattern
  2. Lexi…
     - ➔ T… …graph pattern on t… …ph
  3. GraphFinder
     - ➔ Search the KB graph for the pseudo graph pa…



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Summary

- Three step approach
  1. Graphicator
     - ➔ Turns a natural l… …ery into a ps… pattern
  2. Lexi…
     - ➔ T… …graph pattern on t… …ph …nder…

NLP task

Representational difference needs to be absorbed

LOD task

variation operations

do graph pa…

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Natural Language Interfaces for SPARQL endpoints
- Related Works -

Jin-Dong Kim (DBCLS)

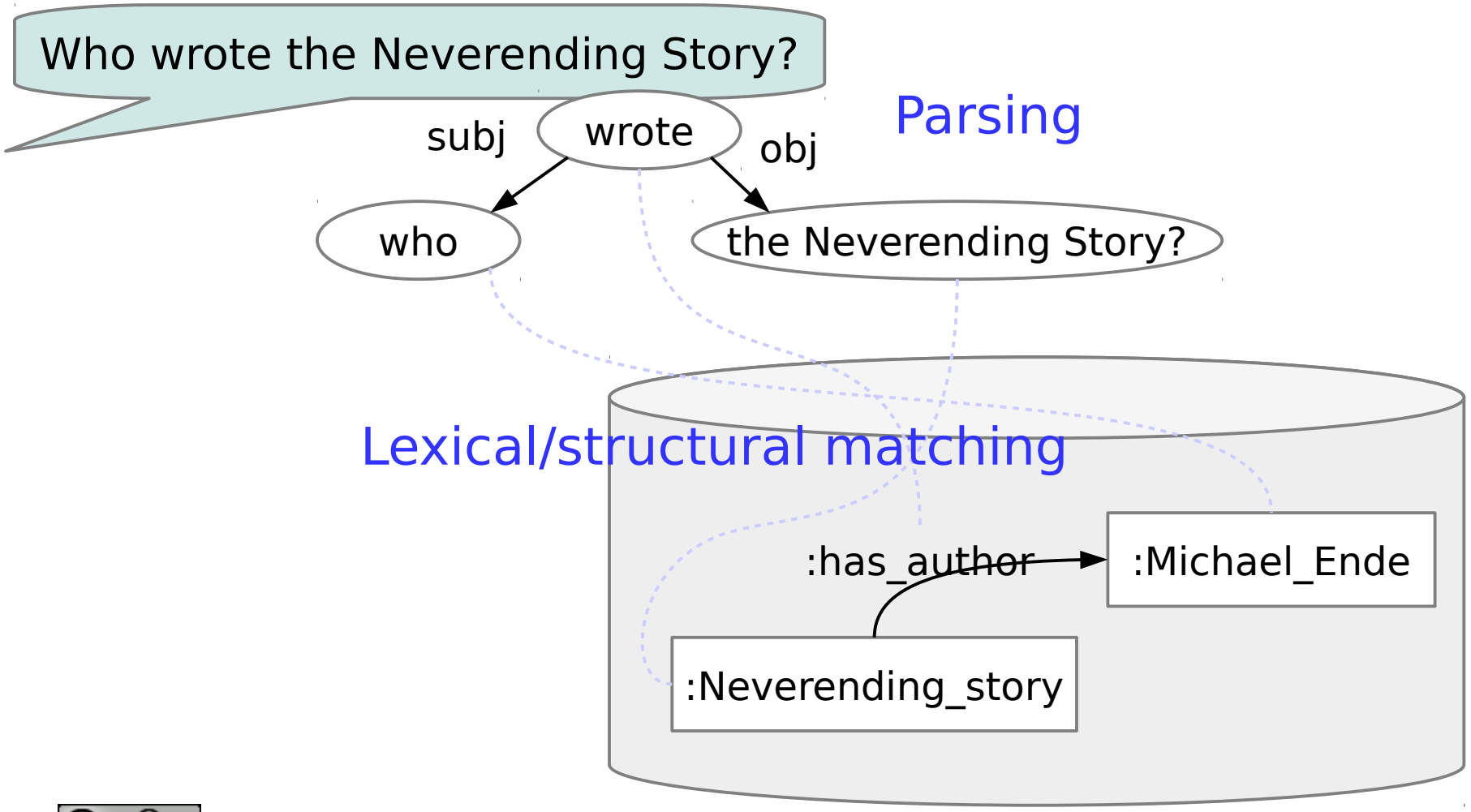
Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Typical approach

- Parsing
- Lexical Matching
- Structural Matching

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Typical approach



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Frontiers

- NQ (2007)
  - ✔ Alexander Ran and Raimondas Lencevicius. 2007. Natural Language Query System for RDF Repositories. In Proceedings of Seventh International Symposium on Natural Language Processing.
- Aqualog (2007)
  - ✔ Vanessa Lopez, Victoria Uren, Enrico Motta, and Michele Pasin. 2007. Aqualog: An ontology-driven question answering system for organizational semantic intranets. Journal of Web Semantics, 5(2):72–105.

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Frontiers

- ORAKEL (2007)
  - ✔ Philipp Cimiano, Peter Haase, and J¨org Heizmann. 2007. Porting natural language interfaces between domains: an experimental user study with the orakel system. In Proceedings of the 12th international conference on Intelligent user interfaces.
- QuestIO (2008)
  - ✔ Valentin Tablan, Danica Damljanovic, and Kalina Bontcheva. 2008. A natural language query interface to structured information. In Proceedings of the 5th European semantic web conference on The semantic web: research and applications.

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Recent systems

- TBQA (AKSW, UManheim, ...)
  - ✔ Template-based SPARQL learner
  - ✔ http://linkedspending.aksw.org/tbsl/
- Treo (DERI)
  - ✔ 'direction' in Gallic
  - ✔ http://treo.deri.de
- LODQA (DBCLS, UColorado, ...)
  - ✔ Linked open data question-answering
  - ✔ http://www.lodqa.org

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

TBSL

- Parsing
  - ✔ LTAG (lexical tree adjoining grammar)
    - ➔ Tree transformation
- Lexical Matching
  - ✔ …
- Structural Matching
  - ✔ **Template generation**

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

TBSL

- To address complex queries
  - ✔ Who produced the most films?
- Generate templates
  - ✔ SELECT ?y WHERE {
    ?x a onto:Film .
    ?x onto:producer ?y
    }
    ORDER BY DESC(COUNT(?x)) OFFSET 0 LIMIT 1

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

TBSL

- To address complex queries
  - ✔ Who produced the most films?
- Generate templates
  - ✔ SELECT ?y WHERE {
    ?x a onto:Film .
    ?x onto:producer ?y
    }
    ORDER BY DESC(COUNT(?x)) OFFSET 0 LIMIT 1

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

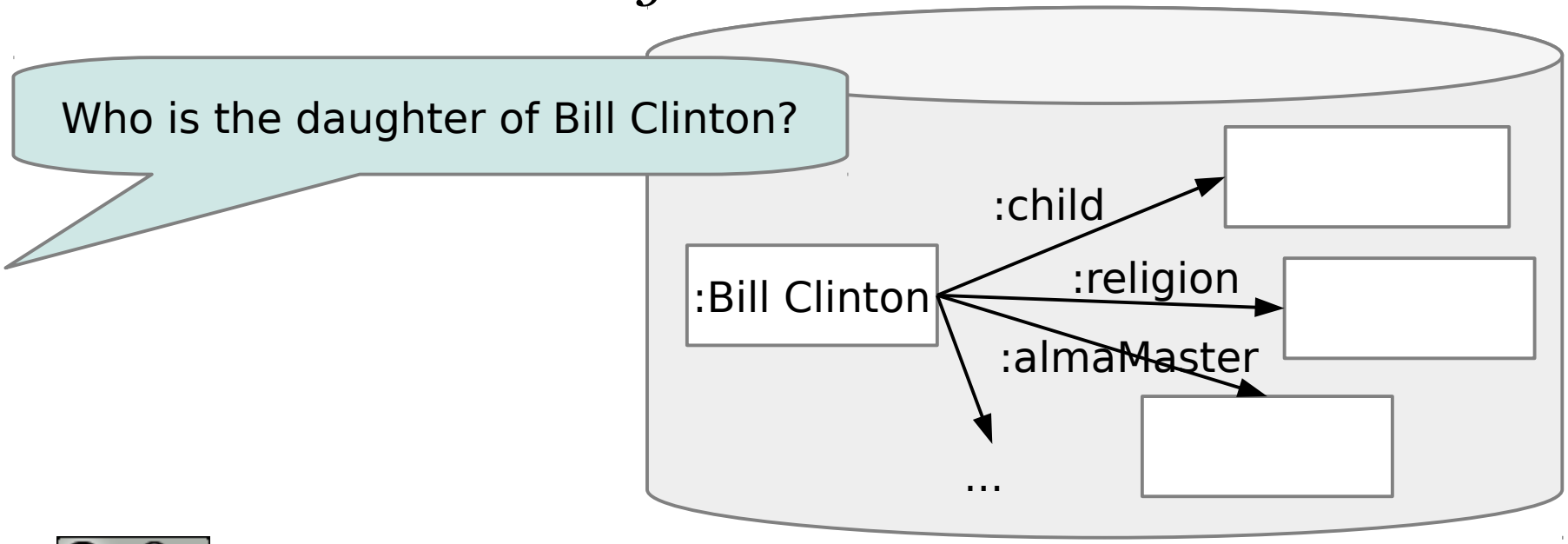
Treo

- Parsing
  - ✔ Dependency parsing
- Lexical Matching
  - ✔ **Distributional semantics**
- Structural Matching
  - ✔ …

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Treo

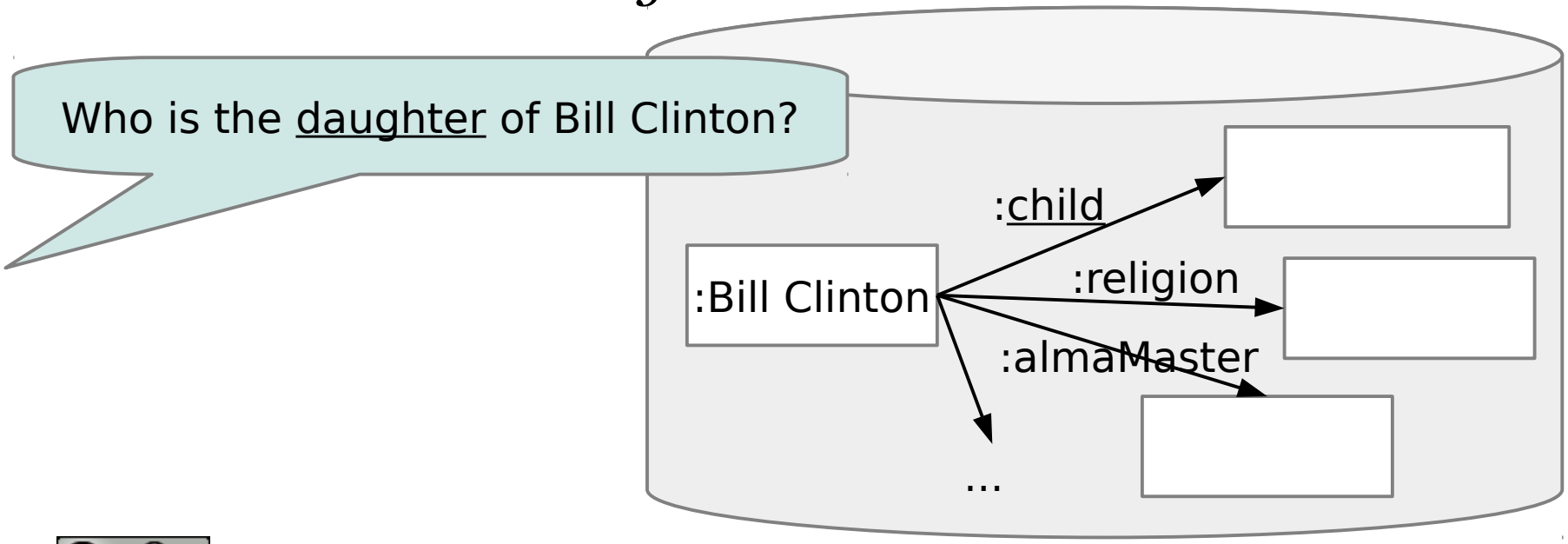
- Lexical matching
  - ✔ Distributional semantics
    - ➔ "*linguistic items with similar distributions have similar meanings.*"



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Treo

- Lexical matching
  - ✔ Distributional semantics
    - ➔ "*linguistic items with similar distributions have similar meanings.*"



Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

LODQA

- Parsing
  - ✔ HPSG (Head-driven Phrasal Structure Grammar)
    - ➔ Graph transformation
- Lexical Matching
  - ✔ …
  - ✔ Public sourcing lexical indexing
- Structural Matching
  - ✔ **Graph variation operations**

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Future directions

Collaborations

- LODQA (DBCLS, UColorado, ...)
  - ✔ Addresses Structural variation problem
- Treo (DERI)
  - ✔ Addresses lexical variation problem
- TBQA (AKSW, UManheim, ...)
  - ✔ Addresses quantifier modeling

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS

Future directions

Collaborations

- LODQA
  - ✔ Invite contribution from the public
    - ➔ Open source
    - ➔ Open more information
    - ➔ Implement more open interface

Licensed under a Creative Commons Attribution 3.0 Unported License - DBCLS