Maverick: Discovering Exceptional Facts from Knowledge Graphs

UNIVERSITY OF TEXAS ARLINGTON

Gensheng Zhang   Damian Jimenez   Chengkai Li

iDiR

Exceptional Facts

abc NEWS — Denzel Washington followed Sidney Poitier as only the second black to win the Best Actor award.

YAHOO! SPORTS — This was Brazil's first own goal in World Cup history.

Chicago Tribune — Hillary Clinton becomes first female presidential nominee.

Entity of interest	Denzel Washington
Context	Best Actor award winners
Attributes	Ethnicity
Peculiar value	African American (only two satisfy)

Given an entity **x** **find** such that the context has many entities, including **x**

A context

A set of attributes (subspace)

x bears a peculiar value w.r.t. the subspace (few in the context have the value)

Applications

Computational Journalism
- Fact-finding
- Fact-checking

E.g., The first female presidential nominee was Victoria Woodhull, not Hillary Clinton (snopes.com)

Data Cleaning

Recommendation Systems
- Friends, news, and product promotion

Willis Tower   Website   Directions

4.4   1,556 Google reviews

Skyscraper in Chicago, Illinois

The Willis Tower, built as and still commonly referred to as Sears Tower, is a 108-story, 1,450-foot skyscraper in Chicago, Illinois, United States. Wikipedia

Hours: Open today · 10AM–3PM ▾

**Did you know:** Willis Tower in Chicago is the second-tallest building in the US. wikipedia.org

Related Work

Outlier detection
- Maverick finds conditions that make an object stand out, although the object may not necessarily be an outlier.

Outlying aspect mining

Challenges in adopting existing algorithms:
- Many assume a single-table model: a graph can be an extremely large and sparse table
- Conjunctive queries on a single table ≠ pattern queries
- Multiple tables: unclear how to handle joins
- Unclear how to handle set values

Exceptional Facts from Knowledge Graphs

What is exceptional about G1? Among all the goals scored by BRA players, G1 is the only own goal.
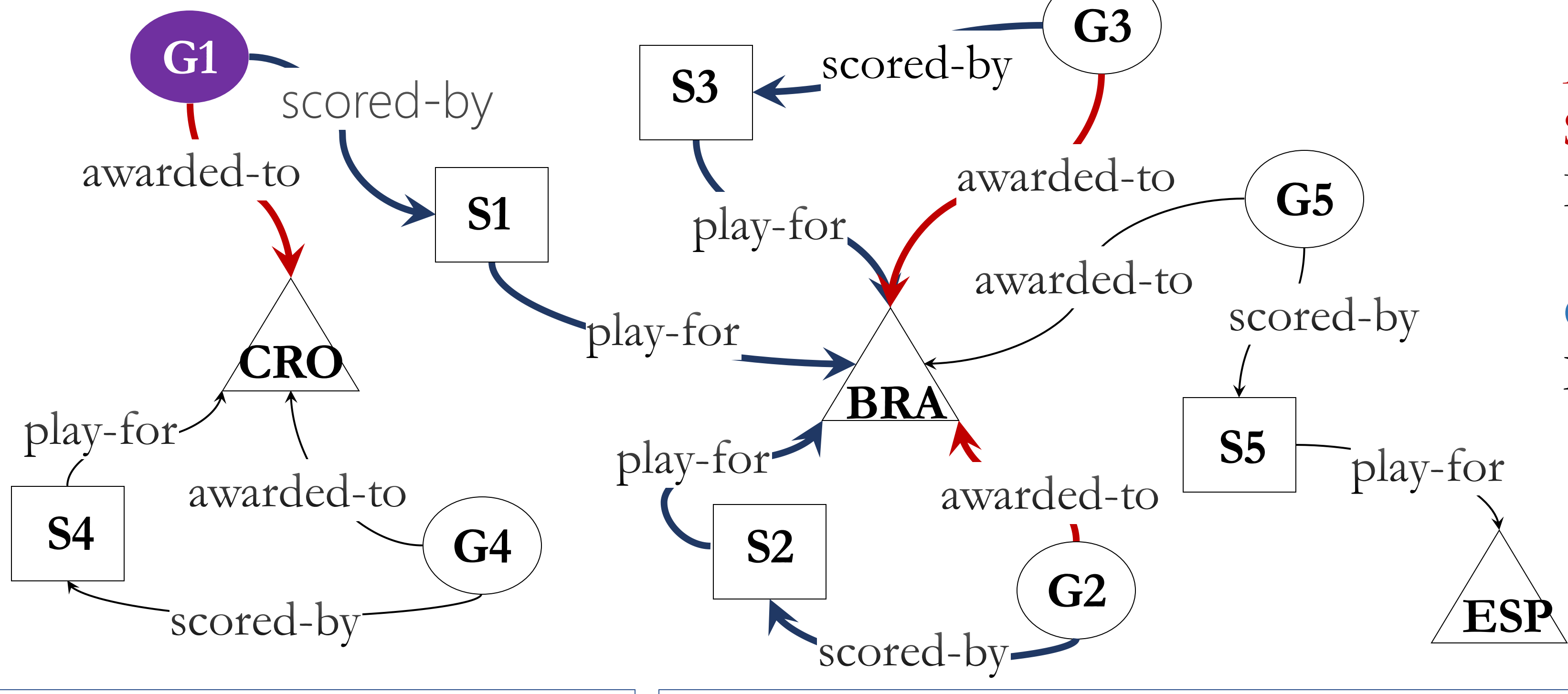
Attributes: labels of incoming/outgoing edges
Subspace: a subset of attributes
E.g., G1. awarded-to = CRO

Context: entities sharing some common characteristics. Defined by a pattern-variable pair
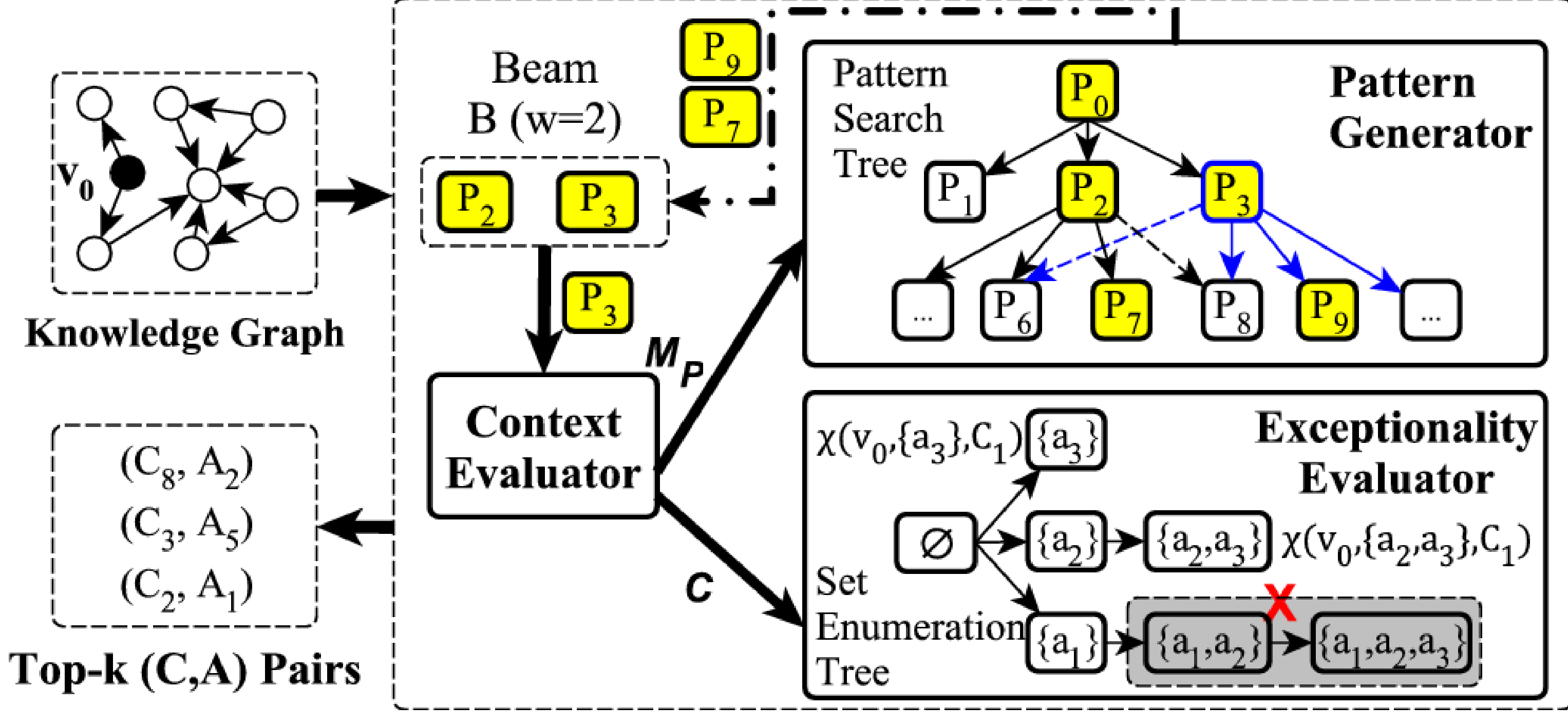
?g —awarded-to→ ?s —play-for→ BRA

E.g., Goals scored by Brazilian players

Problem Formulation

Input
- Entity of interest v_0
- Exceptionality function χ
- Result size k

Output
- Top-k (context, subspace) pairs with regard to χ, in which v_0 stands out

Maverick

Beam B (w=2)

Knowledge Graph

Context Evaluator

(C_8, A_2)
(C_3, A_5)
(C_2, A_1)

Top-k (C,A) Pairs

Pattern Search Tree   **Pattern Generator**

$\chi(v_0,\{a_3\},C_1)\{a_3\}$ **Exceptionality Evaluator**

$\chi(v_0,\{a_2,a_3\},C_1)$
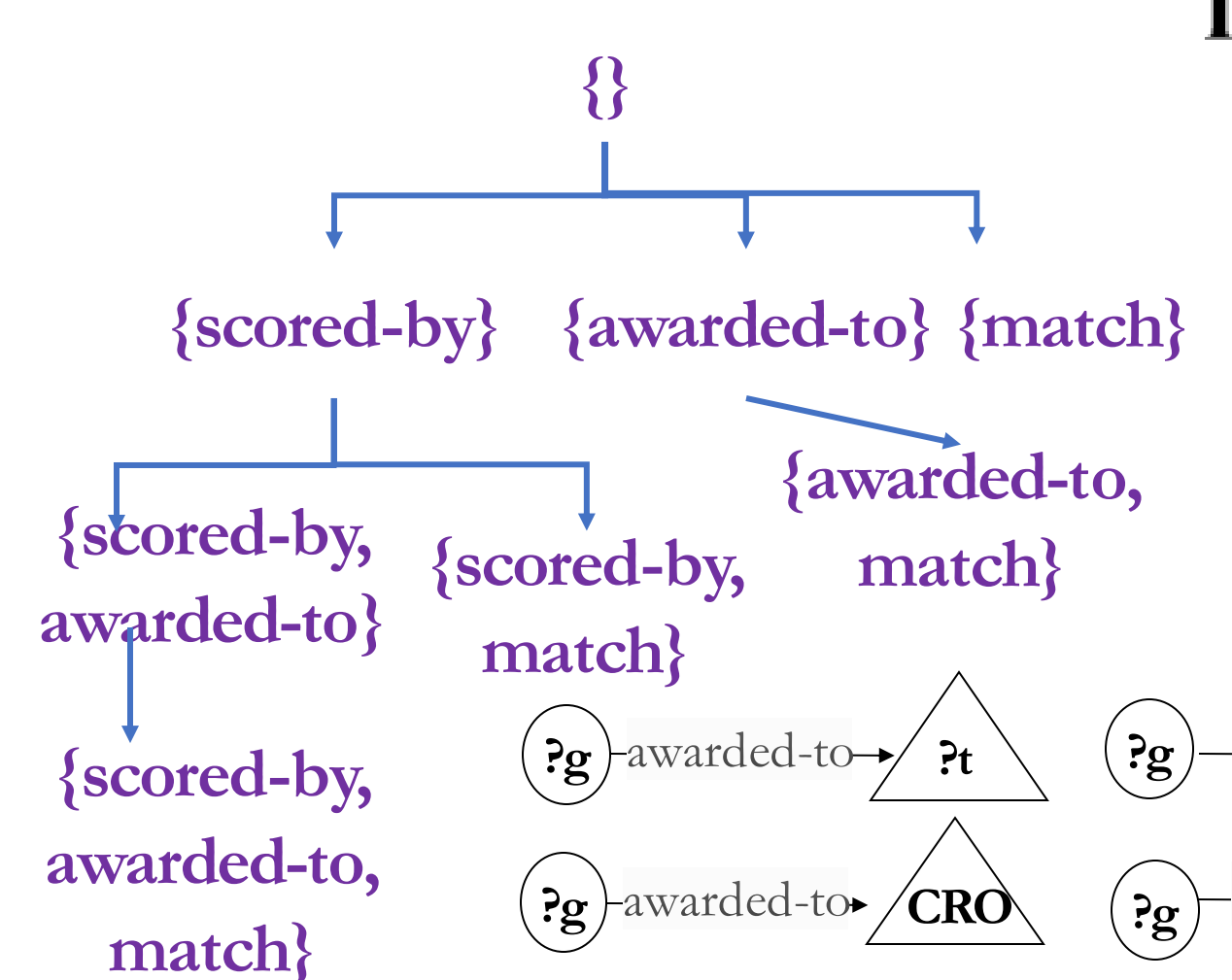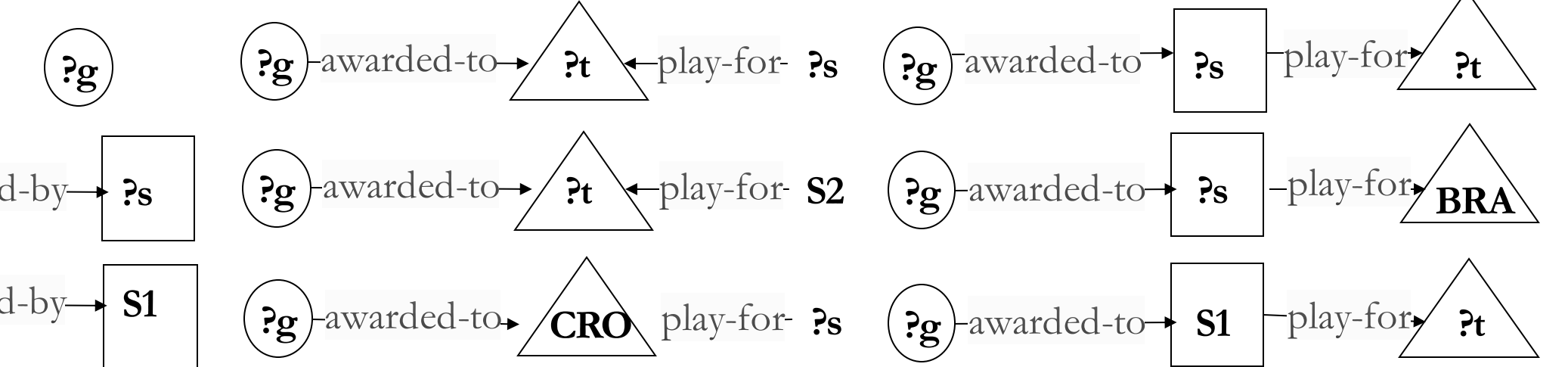
Set Enumeration Tree

Experiments

WCGoals

Created based on FIFA.com
11 node types, 13 edge types
49,078 nodes, 158,114 edges

Film-Award

A subgraph of Freebase
95 node types, 117 edge types
5,437,628 nodes, 10,879,448 edges

Challenges

Number of attribute subspaces: $O(2^{|A_{v_0}|})$
Number of patterns (contexts): $\Omega(2^{|V_G|})$

See you in Rio!
VLDB2018 demo

Beam-Rdm   Beam-Opt   Beam-Conv   Breath-First