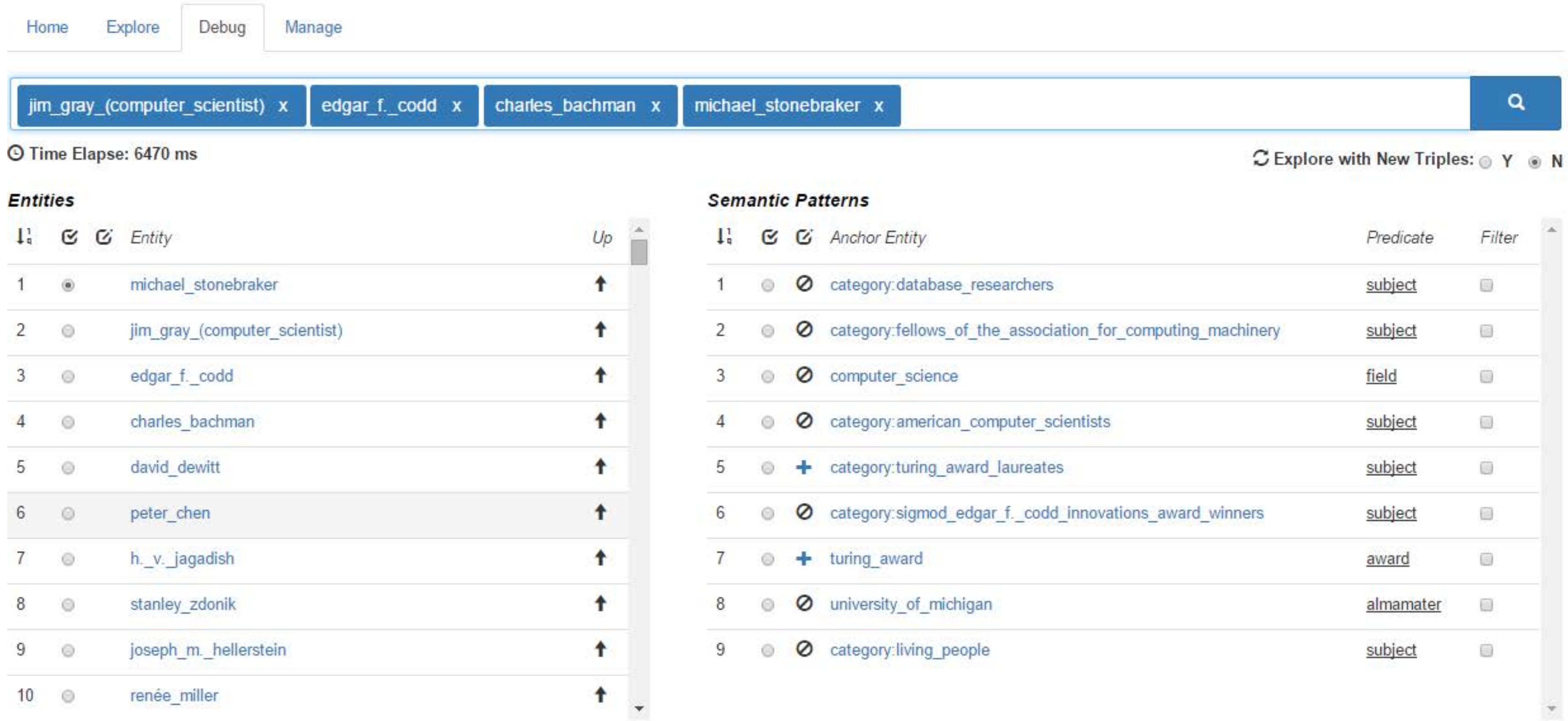
中国人民大学信息学院
SCHOOL OF INFORMATION RENMIN UNIVERSITY OF CHINA

SEED: A System for Entity Exploration and Debugging in Large-scale Knowledge Graph

Jun Chen, Yueguo Chen, Xiaoyong Du, Xiangling Zhang, Xuan Zhou

USER INTERFACE


Home Explore Debug Manage

jim_gray_(computer_scientist) x edgar_f._codd x charles_bachman x michael_stonebraker x

Time Elapse: 6470 ms

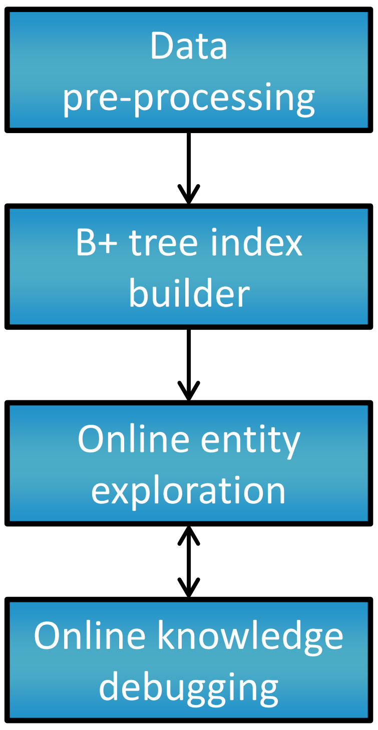
Explore with New Triples: ○ Y ● N

Entities

↓	☑	☑	*Entity*	*Up*
1	●		michael_stonebraker	↑
2	○		jim_gray_(computer_scientist)	↑
3	○		edgar_f._codd	↑
4	○		charles_bachman	↑
5	○		david_dewitt	↑
6	○		peter_chen	↑
7	○		h._v._jagadish	↑
8	○		stanley_zdonik	↑
9	○		joseph_m._hellerstein	↑
10	○		renée_miller	↑

Semantic Patterns

↓	☑	☑	*Anchor Entity*	*Predicate*	*Filter*
1	○	⊘	category:database_researchers	subject	□
2	○	⊘	category:fellows_of_the_association_for_computing_machinery	subject	□
3	○	⊘	computer_science	field	□
4	○	⊘	category:american_computer_scientists	subject	□
5	○	+	category:turing_award_laureates	subject	□
6	○	⊘	category:sigmod_edgar_f._codd_innovations_award_winners	subject	□
7	○	+	turing_award	award	□
8	○	⊘	university_of_michigan	almamater	□
9	○	⊘	category:living_people	subject	□


Data pre-processing

B+ tree index builder

Online entity exploration

Online knowledge debugging

Entity-oriented

Interactive

Visual

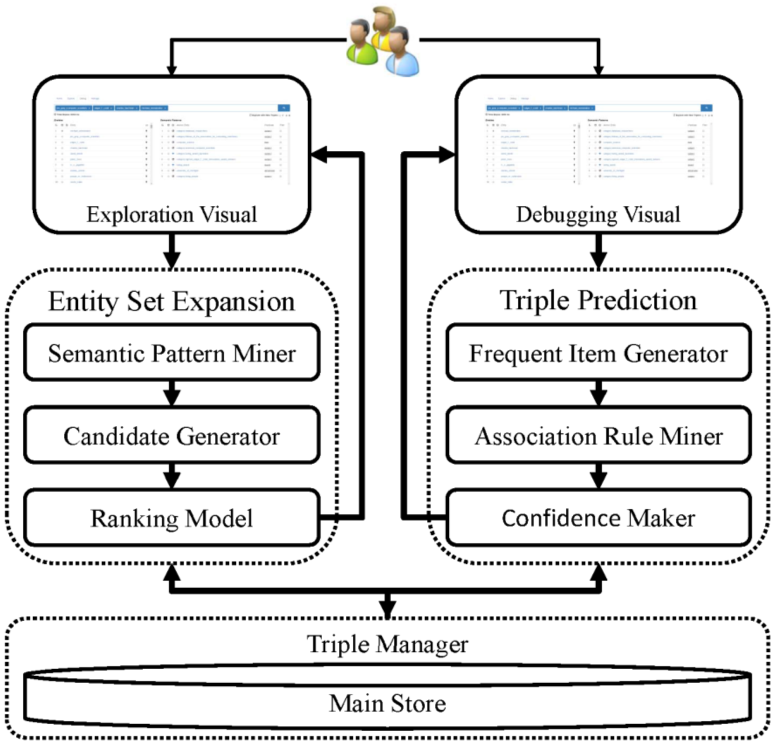
Semantic

INTRODUCTION

SEED is built on large-scale knowledge graphs (KGs). It takes entities as inputs, and outputs similar entities and their semantic patterns. It is designed to assist users exploring the KGs with two key functions:

1. Entity exploration;
2. Knowledge debugging.

SYSTEM ARCHITECTURE


Exploration Visual

Debugging Visual

Entity Set Expansion

Semantic Pattern Miner

Candidate Generator

Ranking Model

Triple Prediction

Frequent Item Generator

Association Rule Miner

Confidence Maker

Triple Manager

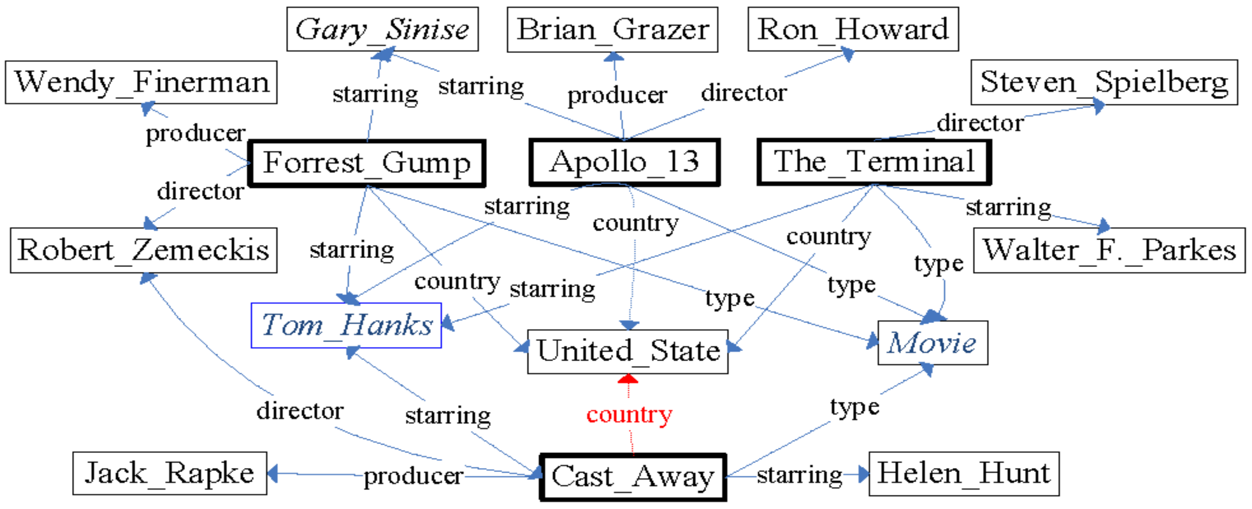
Main Store

KEY TECHNIQUES

Entity set expansion: Given several seed entities (Q), e.g. *Forrest Gump*, *Apollo 13*, *The Terminal*. Firstly, we discover semantic patterns (SPs) among them, e.g. *starring-Tom Hanks*. Secondly, we expand candidate entities satisfying the discovered SPs, e.g. *Cast Away*. Finally, we rank candidate entities (e) evaluating their similarity to Q as following:

$$\mathrm{Similarity}(e,Q) = \sum_{(SP\in e'|e'\in Q)\wedge SP\in e} \mathrm{Relevance}(SP,Q) * \mathrm{Discriminability}(SP)$$

Triple prediction: Given a triple consisting an entity and a SP not existing in KGs, e.g. *Cast Away-country-United State*. Firstly, we mine frequent items of the entities satisfying the specific SP named *country-United State*. Secondly, we generate association rules based on above frequent items. Finally, we recommend the confidence degree via evaluating the likelihood of the missing triple.



Demo URL: 222.29.197.244:8080/SEED
Jun Chen (chenjun2013@ruc.edu.cn)