1. Motivation

A business analyst is investigating the development of Silicon Valley. Companies in Silicon Valley? Persons Graduated from Stanford?

Google
HP
Yahoo!

Jerry Yang
Larry Page
David Packard

David Filo

2. Pain with Search Engine

Search silicon valley company

for each

Search Yahoo! founder

for each

Search David Filo stanford graduate

<Jerry Yang, Yahoo!> confirmed to be a right answer

3. Entity-Relationship Query

An entity-centric, declarative, structured, keyword-based query mechanism.

SELECT x, y
FROM PERSON AS x, COMPANY AS y
WHERE x:[“Stanford” “graduate”] ..... p1
AND y:[“Silicon Valley”] ........... p2
AND x,y:[“found”] .............. p3

Typed entity Variable
Relation Predicate
Selection Predicate

4. Query Answer

Co-occurrence contexts as evidence

For p1: Stanford graduates William Hewlett and David Packard ...

For p2: Hewlett-Packard, in Silicon Valley, ...

For p3: William Hewlett and David Packard founded Hewlett-Packard.

<David Packard, HP> is an answer

5. System

Offline Components

Preprocessor
- cleaning, stemming, sentence segmentation

Indexer
- assign entities to types based on categories

Entity Classifier
- assign entities to types based on categories

Online Components

Query Interface

Ranker
- retrieve entities and evidence

Retriever
- retrieve entities and evidence


- 2 million Wikipedia articles
- 0.75 million entities
- 10 predefined entities
- COMPANY, NOVEL, PERSON ...
- 100 million annotations
- INEX17
- Queries adapted from INEX topics
- OWN28
- Our own crafted queries

Pride and Prejudice

Pride and Prejudice is a novel by Jane Austen. It was begun in 1796; it was her second attempt ...

Categories: British novels

7. Ranking

Bounded Cumulative Model

\[ r(t) = \sum_{o \in O} f(o) \left[ 1 - \prod_{s \in \mathcal{S}} \left( 1 - \text{prox}(t, s) \cdot \text{cred}(o, s) \right) \right] \]

Ordering Pattern
- Frequent patterns indicate more reliable evidence.

Proximity
- High proximity indicates more reliable evidence.

Mutual Exclusion
- In a context with colliding patterns, the pattern followed by more prominent entities is more likely to be effective.
