EntityEngine: Answering Entity-Relationship Queries using Shallow Semantics

http://idir.uta.edu/eq

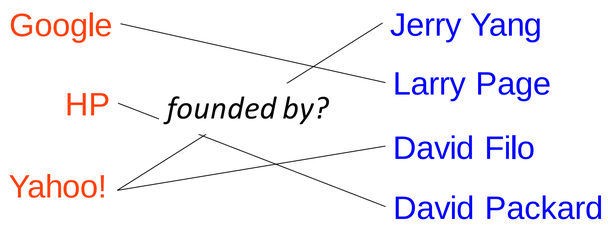
Xiaonan Li Chengkai Li
University of Texas at Arlington

Cong Yu
Yahoo! Research

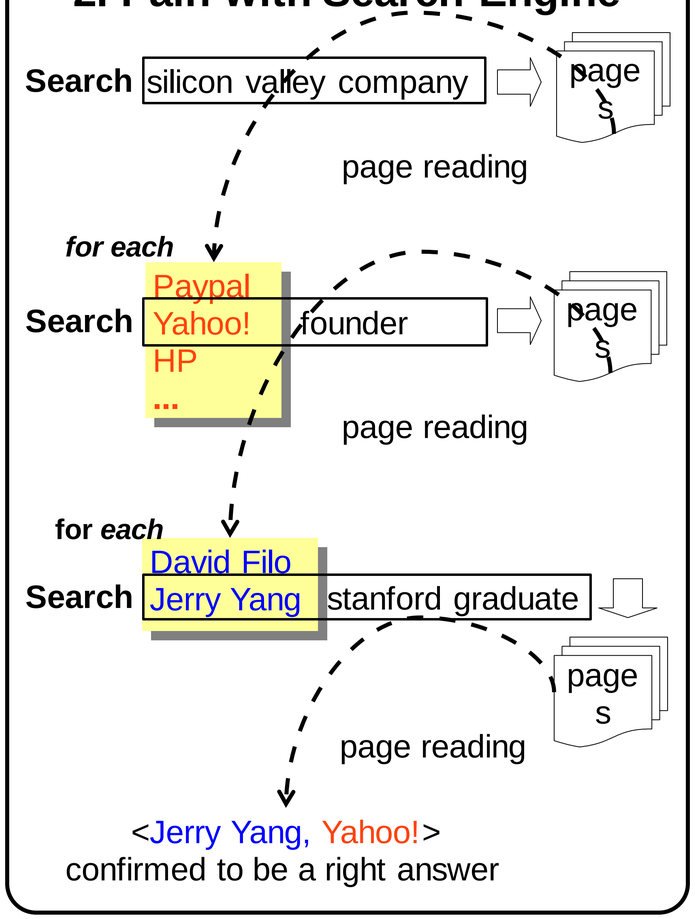
1. Motivation

A business analyst is investigating the development of Silicon Valley.

Companies in Silicon Valley? — *Persons Graduated from Stanford?*



2. Pain with Search Engine



3. Entity-Relationship Query

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

Typed entity Variable

```
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
```

Relation Predicate

Selection Predicate

4. Query Answer

Co-occurrence contexts as **evidence**

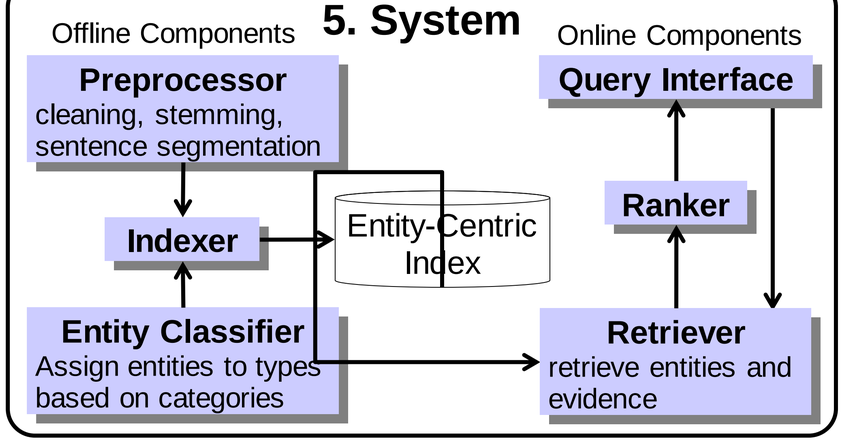
For p1: Stanford graduates William Hewlett and David Packard …

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



6. Demonstration on Wikipedia

- 2 million Wikipedia articles
- 0.75 million entities
- 10 predefined
  - 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

Pride and Prejudice is an entity of type NOVEL.
It is also an article, in which the PERSON Jane Austen occurs.

7. Ranking

answer	x	y	p1	p2	p3	Ranking score
t1	Jerry Yang	Yahoo!	0.8	0.7	0.8	**0.448**
t2	Larry Page	Google	0.6	0.5	0.6	**0.180**
t3	Scott McNealy	Cisco	0.9	0.8	0.2	**0.144**
t4	Bill Gates	IKEA	0.3	0.1	0.2	**0.006**

0.8 * 0.7 * 0.8 = 0.448

Bounded Cumulative Model

$$F_p(t) = \sum_{o \in O_p} f(o)[1 - \prod_{s \in \Phi_p(t,o)}(1 - prox(t,s)\,credit(o,s))]$$

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.

[1] EntityEngine: Answering Entity-Relationship Queries using Shallow Semantics. In *CIKM* (demo), 2010.

[2] Entity-Relationship Query over Wikipedia. In *SMUC Workshop* (in conjunction with *CIKM*), 2010.