TEL AVIV UNIVERSITY אוניברסיטת תל-אביב

Penn* University of Pennsylvania

Ontology Assisted Crowd Mining

Yael Amsterdamer **Susan B. Davidson*** **Tova Milo**

Slava Novgorodov **Amit Somech**

Goals

- Crowd data sourcing is a powerful data procurement paradigm
- We would like harness it in order to allow users to:
  - specify their information needs in a declarative manner
  - efficiently mine the crowd for relevant data
  - obtain a concise list of answers that represent frequent, significant data patterns

A Motivating Example

Ann is planning a vacation in NYC with her family:

"I'm looking for activities to do at a child-friendly attraction in New York, and a good restaurant near by"

Answers should be relevant, represent popular recommendations, and may include additional advice from the crowd

"You can go bike riding in Central Park and eat at Maoz Vegetarian.
Tips: Rent bikes at the boathouse"

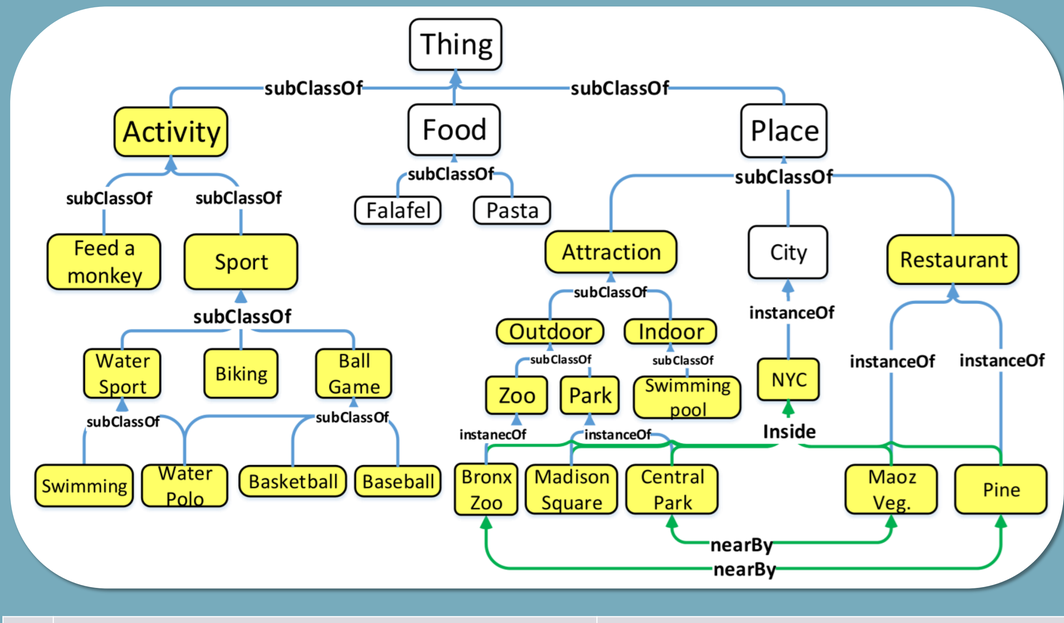
"You can go visit the Bronx Zoo and eat at Pine Restaurant.
Tips: Order antipasti at Pine.
Skip dessert and go for ice cream across the street"

Web search can only extract information from recorded data
- It may be hard to extract combinations of restaurant and activity

Forums can yield targeted answers,
- which require reading, aggregating, identifying consensus, etc.

We propose a new, alternative approach using crowd mining

Formal Model Based on RDF



T1	I visited the Bronx Zoo and ate pasta at Pine on April 5th	[Visit doAt Bronx_Zoo]. [Pasta eatAt Pine]
T2	I played basketball in Central Park on April 13th	[Basketball playAt Central_Park]
T3	I played baseball in Central Park and ate falafel at Maoz Veg. on April 27th	[Baseball playAt Central_Park]. [Falafel eatAt Maoz_Veg]
	…	…

OASSIS-QL: The Mining Language

```
SELECT VARIABLES
WHERE
    {$w subClassOf* Attraction
     $x instanceOf $w.
     $x inside      NYC.
     $y subClassOf* Activity.
     $z instanceOf Restaurant.
     $z nearBy      $x}
SATISFYING
     {$y+ doAt      $x.
     []  eatAt      $z.
     MORE}
     WITH SUPPORT = 0.03
```

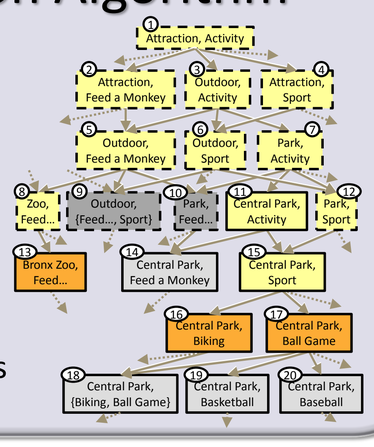
Evaluated over the ontology, to identify candidate data patterns

Retain the patterns that are significant for the crowd, and find additional advice

Efficient Query Evaluation Algorithm

- Lazily construct a semantic subsumption partial order
- Traverse it in a top-down manner
- Automatically generate crowd questions
- Prune insignificant parts

The goal is minimizing the # crowd questions



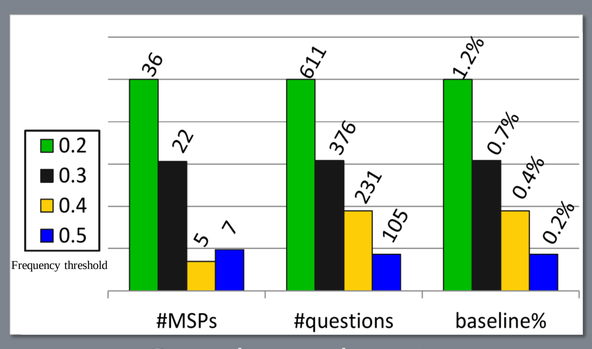
Additional Aspects of the Algorithm

- Open questions – letting crowd members specify patterns

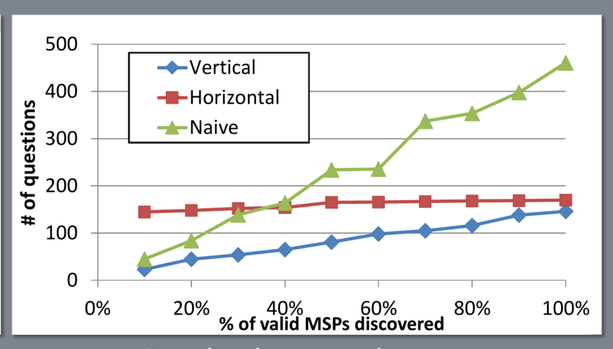
"What else do you do when you play basketball in Central Park?"

- Asking a sequence of questions "in context"
- Quick pruning of irrelevant items by crowd members
- Multiple crowd workers in parallel
- Output quality assurance

Experimental Results



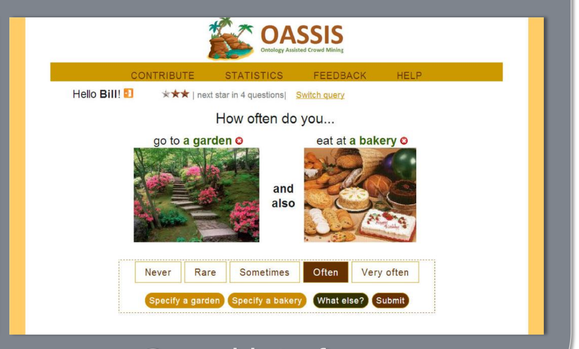
Crowd experiments



Synthetic experiments

OASSIS Screenshots

Query builder



Crowd interface