

Micro-task Crowdsourcing Applications for Entity Linking and Query Understanding

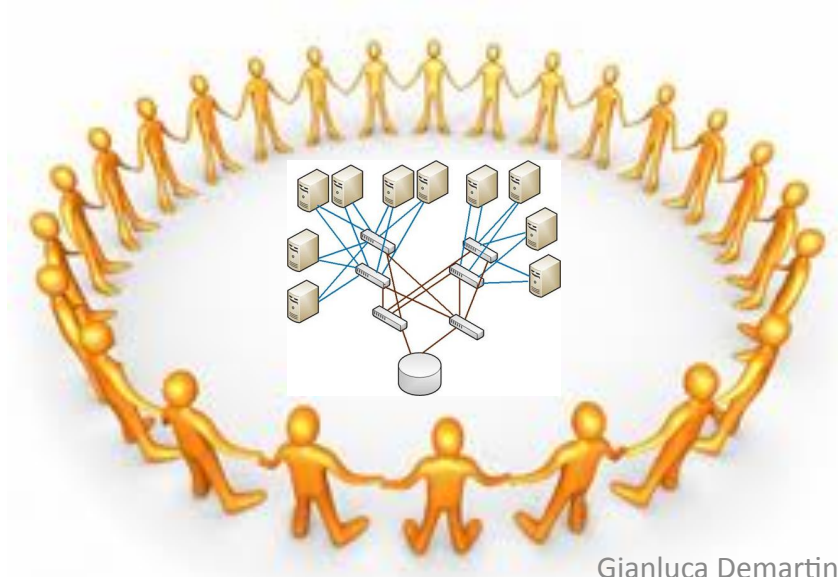
Gianluca Demartini

gianlucademartini.net

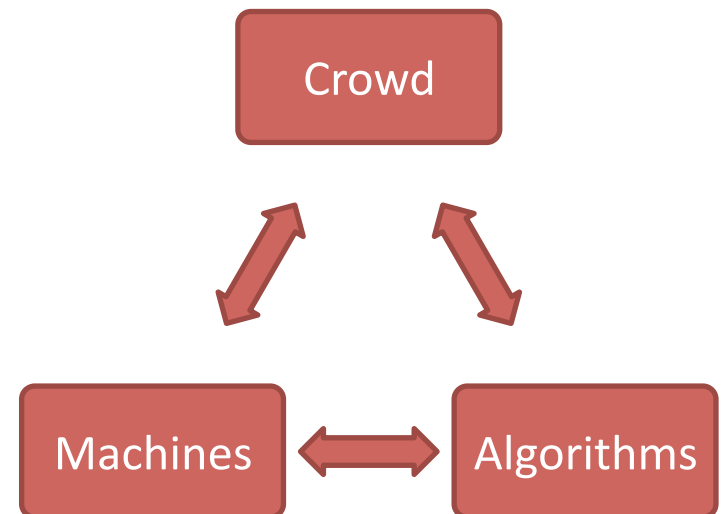
exascale.info

ZenCrowd: Entity Linking by the Crowd

- Combine both algorithmic and manual linking
- Automate manual linking via crowdsourcing
- Dynamically assess human workers with a probabilistic reasoning framework



Gianluca Demartini



Facebook Buys Instagram for \$1 Billion

BY EVELYN M. RUSLI

2:02 p.m. | Updated

Facebook is not waiting for its initial public offering to make its first big purchase.

In its largest acquisition to date, the social network has purchased Instagram the popular photo-sharing application, for about \$1 billion in cash and stock, the company said Monday.



<http://dbpedia.org/resource/Facebook>

HTML:

<p>Facebook is not waiting for its initial public offering to make its first big purchase.</p><p>In its largest acquisition to date, the social network has purchased Instagram, the popular photo-sharing application, for about \$1 billion in cash and stock, the company said Monday.</p>

<http://dbpedia.org/resource/Instagram>

owl:sameAs

fbase:Instagram

RDFa
enrichment

<p><cite property="rdfs:label">Facebook</cite> is not waiting for its initial public offering to make its first big purchase.</p><p>In its largest acquisition to date, the social network has purchased <cite property="rdfs:label">Instagram</cite>, the popular photo-sharing application, for about \$1 billion in cash and stock, the company said Monday.</p>

CNET > News > Mobile

Instagram for Android is now available

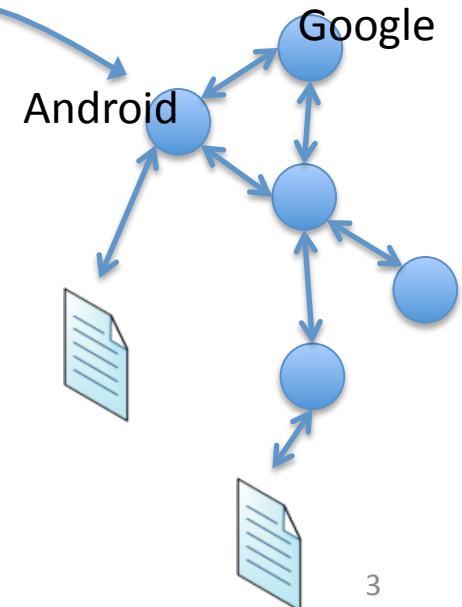
At long last, Instagram finally releases the Android version of its app.



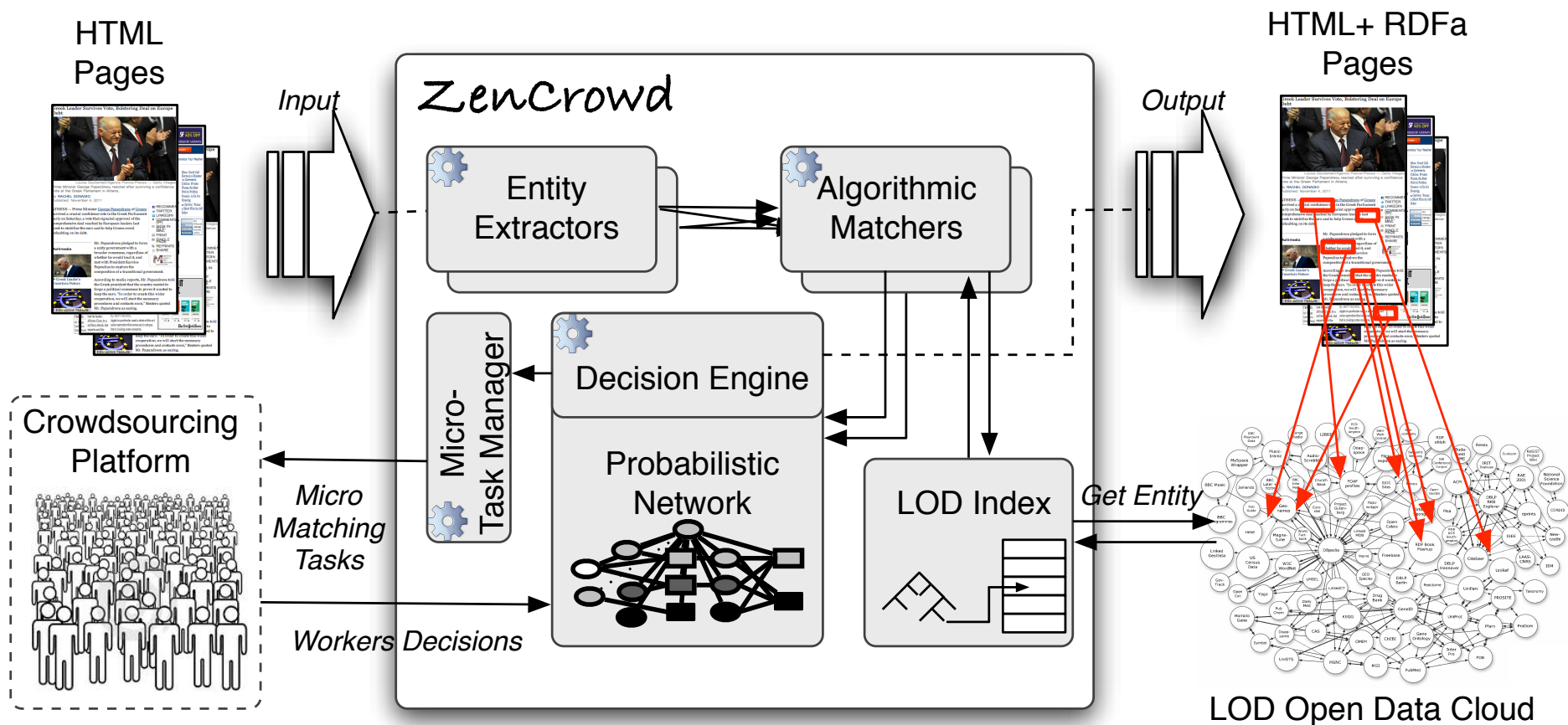
by Jason Cipriani | April 3, 2012 10:07 AM PDT

Follow

Instagram has been around since 2010, available only to iOS devices. Android users have been waiting patiently, with repeated promises of an Android version arriving soon.



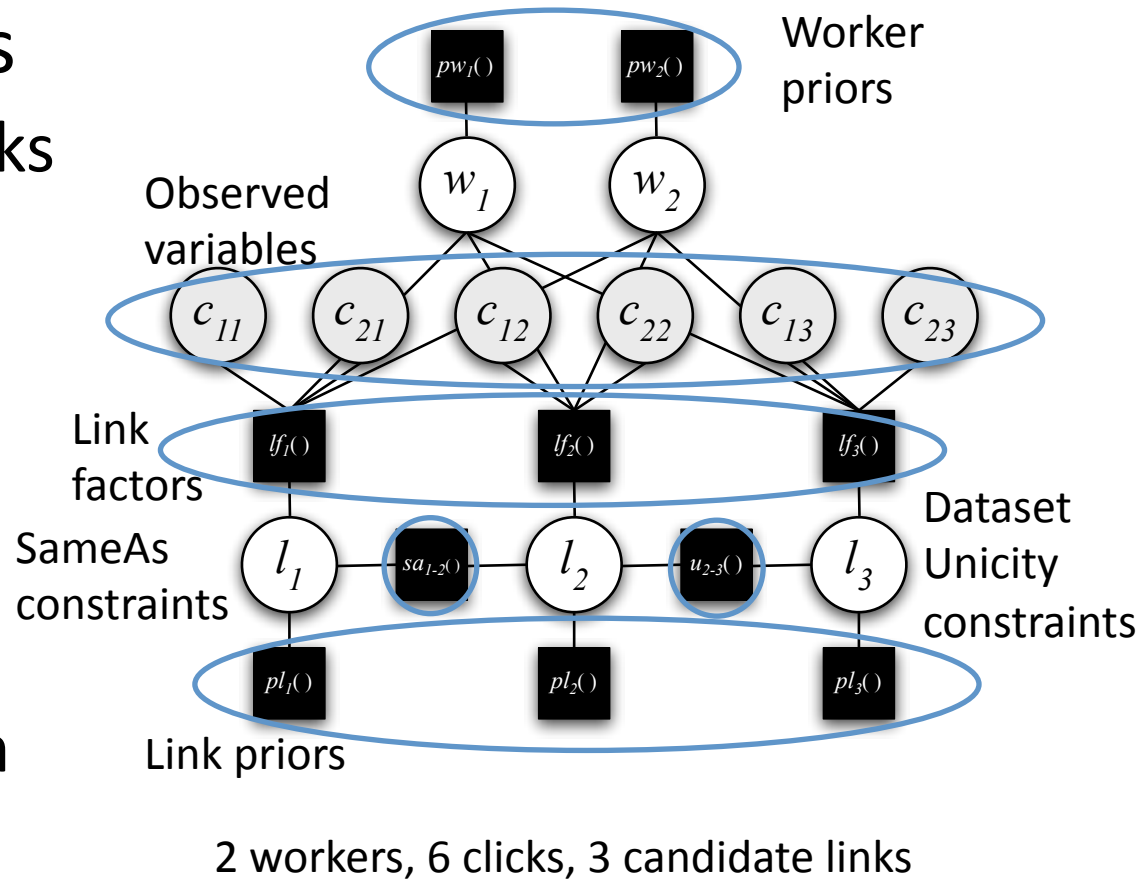
ZenCrowd Architecture



Gianluca Demartini, Djellel Eddine Difallah, and Philippe Cudré-Mauroux. ZenCrowd: Leveraging Probabilistic Reasoning and Crowdsourcing Techniques for Large-Scale Entity Linking. In: 21st International Conference on World Wide Web (WWW 2012).

Entity Factor Graphs

- Graph components
 - Workers, links, clicks
 - Prior probabilities
 - Link Factors
 - Constraints
- Probabilistic Inference
 - Select all links with posterior prob $> \tau$



Lessons Learnt

- Crowdsourcing + Prob reasoning works!
- But
 - Different worker communities perform differently
 - Many low quality workers
 - Completion time may vary (based on reward)
- Need to find the right workers for your task (see WWW13 paper)

ZenCrowd Summary

- ZenCrowd: Probabilistic reasoning over automatic and crowdsourcing methods for entity linking
- Standard crowdsourcing improves 6% over automatic
- 4% - 35% improvement over standard crowdsourcing
- 14% average improvement over automatic approaches

<http://exascale.info/zencrowd/>

- Follow up-work (VLDBJ):
 - Also used for instance matching across datasets
 - 3-way blocking with the crowd

CrowdQ – Crowd-powered Query Understanding

Motivation

- Web Search Engines can answer simple factual queries directly on the result page
- Users with complex information needs are often unsatisfied
- Purely automatic techniques are not enough
- We want to solve it with Crowdsourcing!

CrowdQ

- CrowdQ is the first system that uses crowdsourcing to
 - *Understand* the intended meaning
 - *Build* a structured query template
 - *Answer* the query over Linked Open Data

Gianluca Demartini, Beth Trushkowsky, Tim Kraska, and Michael Franklin. CrowdQ: Crowdsourced Query Understanding. In: 6th Biennial Conference on Innovative Data Systems Research (CIDR 2013).

birthdate of the mayors of all the cities in Italy



About 124,000,000 results (0.33 seconds)

City	Mayor	Birthdate
Rome, Italy	Gianni Alemanno	March 3, 1958
Venice, Italy	Giorgio Orsoni	August 29, 1946
Milan, Italy	Giuliano Pisapia	May 20, 1949

[Press to see more](#)

[Cities in Italy | Italy Travel Guide](#)

www.italylogue.com/italian-cities

Learn about the best **cities in Italy** to visit, and some **Italian cities** you might never have heard of before. These **cities in Italy** are **all** great for visitors.

[Top Ten Cities for Visitors to Italy - Top Italian Cities to See](#)

goitaly.about.com/od/planningandinformation/tp/topcities.htm

Italy has many beautiful and historic **cities** that are well worth a visit. Here are our picks for the ten best **cities** for visitors to **Italy**.

[Italian Cities and Towns - Italy](#)

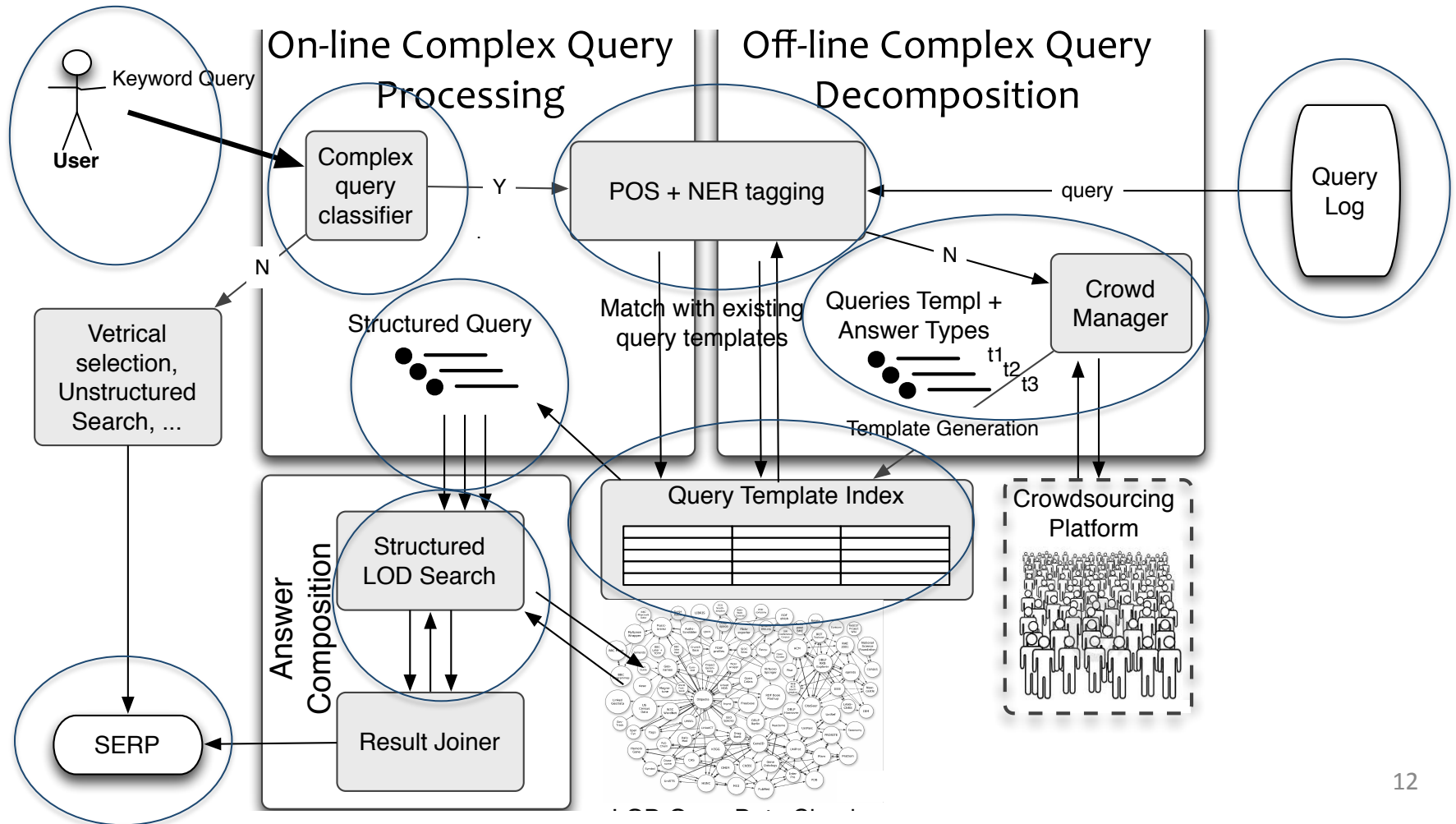
en.comuni-italiani.it/

Information and statistics on **Italian** Regions, Provinces, and Municipalities. **All Cities**

CrowdQ Architecture

Off-line: query template generation with the help of the crowd

On-line: query template matching using NLP and search over open data



Hybrid Human-Machine Pipeline

Q= birthdate of actors of forrest gump

Query annotation

Noun

Noun

Named entity

Verification

Is forrest gump this entity in the query?

Entity Relations

Which is the relation between: actors and forrest gump → starring

Schema element

Starring → <dbpedia-owl:starring>

Verification

Is the relation between:
Indiana Jones – Harrison Ford
Back to the Future – Michael J. Fox
of the same type as
Forrest Gump - actors

Structured query generation

Q= birthdate of actors of fo

MOVIE

SELECT ?y ?x

WHERE { ?y <dbpedia-owl:birthdate> ?x .

?z <dbpedia-owl:starring> ?y .

?z <rdfs:label> 'Fo

MOVIE

mp' }

Results from BTC09:

```
<http://dbpedia.org/resource/Robin_Wright_Penn> 1966-04-08
<http://dbpedia.org/resource/Tom_Hanks> 1956-07-09
<http://dbpedia.org/resource/Sally_Field> 1946-11-06
<http://dbpedia.org/resource/Gary_Sinise> 1955-03-17
<http://dbpedia.org/resource/Mykelti_Williamson> 1960-03-04
```