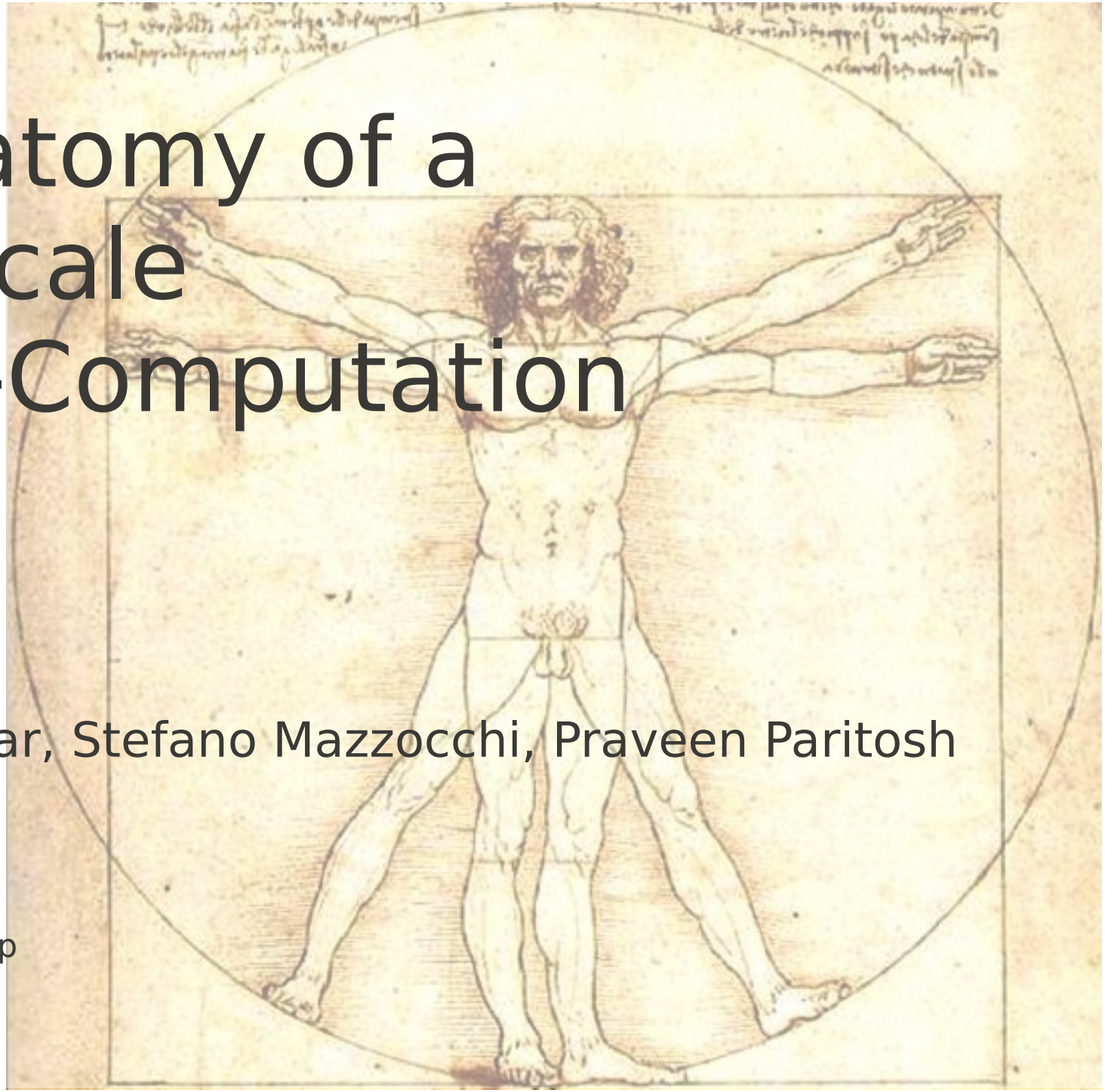
The Anatomy of a Large-Scale Human-Computation Engine



Shailesh Kochhar, Stefano Mazzocchi, Praveen Paritosh

Freebase August Meetup

1: Freebase & Human Computation

2: Example – Stanford Library

3: RABJ

4: Consensus

Freebase

Structured database

12 MM entites, 300 MM triples/facts

Where does the data come from?

Community contributions

Mass Data Loads

Human Judgments Improve Both

Community

Simplify contribution through games

TYPEWRITER

Signed in as: kochhar | Sign Out

The Freebase Data Team made some guesses about which topics belong in these types. Choose a type below and for each topic click Yes if you think we guessed or No if you think we got it wrong. You'll learn something new, share what you already know with the Freebase community and earn contribution points in doing so!

Typewriter » Ethnicity

Undo

Is this an Ethnicity?

from: *Ethnic groups in Tanzania*

Yes Skip No

Gogo people

The Gogo (or Wagogo or Chigogo) are a Bantu ethnic and linguistic group based in the Dodoma Region of central Tanzania. In 1992 the Gogo population was estimated to number 1,300,000. The Gogo have historically been predominantly pastoralist and patrilineal (tracing descent and inheritance through the male line), but many contemporary Gogo now practice settled agriculture, have migrated to urban areas, or work on plantations throughout Tanzania. Gogo music has achieved an international reputation. Their name was invented sometime in the 19th century by the Nyamwezi caravans passing through the area while it was still frontier territory. Richard Francis Burton claimed a very small population for it, saying only that a person could walk for two weeks and find only scattered Tembes. There was and remains the problem of inadequate rain for crops and humans, the rainy season being short and erratic with frequent drought. In the 18th century the Wagogo were mostly pioneer colonists from...

permalink

from: *Ethnic groups in Africa*

Yes Skip No

Tajakant

The Tajakant (also Tadjakant) is a Sahrawi-Moorish tribe of Arab-Berber origins. They speak Hassaniya Arabic. They traditionally lived in Western Sahara, Mauritania and Algeria. They are Muslims, belonging to the Maliki school of Sunni Islam. Some sources claim they are

LEADERBOARD

THIS WEEK		91
1	hangy	63
2	tanmaydalvi	14
2	tfmorris	14
	kochhar	**0**

ALL TIME		1,090,641
1	kjwcode	120,856
2	ranahassan	68,766
3	reaj	54,807
4	sprocketonline	51,086
5	doconnor	46,570
6	tfmorris	45,127
7	akeko	44,710
8	jon	41,531
9	gangofnuns	39,267
10	krsalis	21,904
101	**kochhar**	**650**

http://typewriter.freebaseapps.com/

Community

Simplify contribution through games

Enable QA for Gridworks loads

The For Real Filmfare Best Movie Awards by jon fb

Started: a week ago
Finished: a week ago

▶ Show Load Data

Queue	Answers		Judgements		Score
Gridworks QA: Producer 1 (gridworks/115)		155/155		332/310	▼ 93.5%

Discording (10) | Uncertain (0) | Concording (145)

1. **reconciled to /en/mukesh_bhatt [question_6e3b5a8eddbc9703]**
   - new - mikan
   - skip - jmontilla
   - reconciled to /en/mukesh_bhatt - mai_1980
   - reconciled to /en/mukesh_bhatt - eocay

   ▶ Show

2. **new [question_7dc89aa97fd5532b]**
   - reconciled to /m/0clbdm6 - dfhuynh
   - reconciled to /m/0clbdm6 - eocay

   ▶ Show

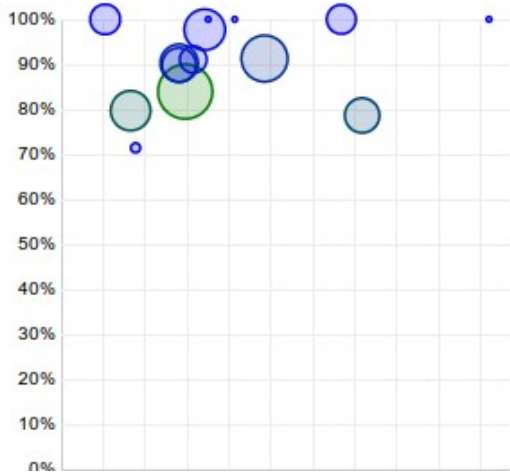
3. **reconciled to /en/nazir_ahmed [question_6bd1dbf390d1d067]**
   - new - eocay
   - reconciled to /en/nazir_ahmed - mai_1980
   - skip - jmontilla
   - skip - mikan

   ▶ Show

4. **reconciled to /en/mukesh_bhatt [question_0d4cca916effd364]**
   - reconciled to /en/mahesh_bhatt - dfhuynh
   - reconciled to /en/mukesh_bhatt - mai_1980

Gridworks QA: Production Company 1 (gridworks/115)	117/117	238/234	▶ 98.3%
Gridworks QA: Film2 (gridworks/115)	225/225	479/450	▶ 91.1%
Gridworks QA: Producer 2 (gridworks/115)	35/35	70/70	100.0%
Gridworks QA: Production Company 2 (gridworks/115)	3/3	9/6	100.0%

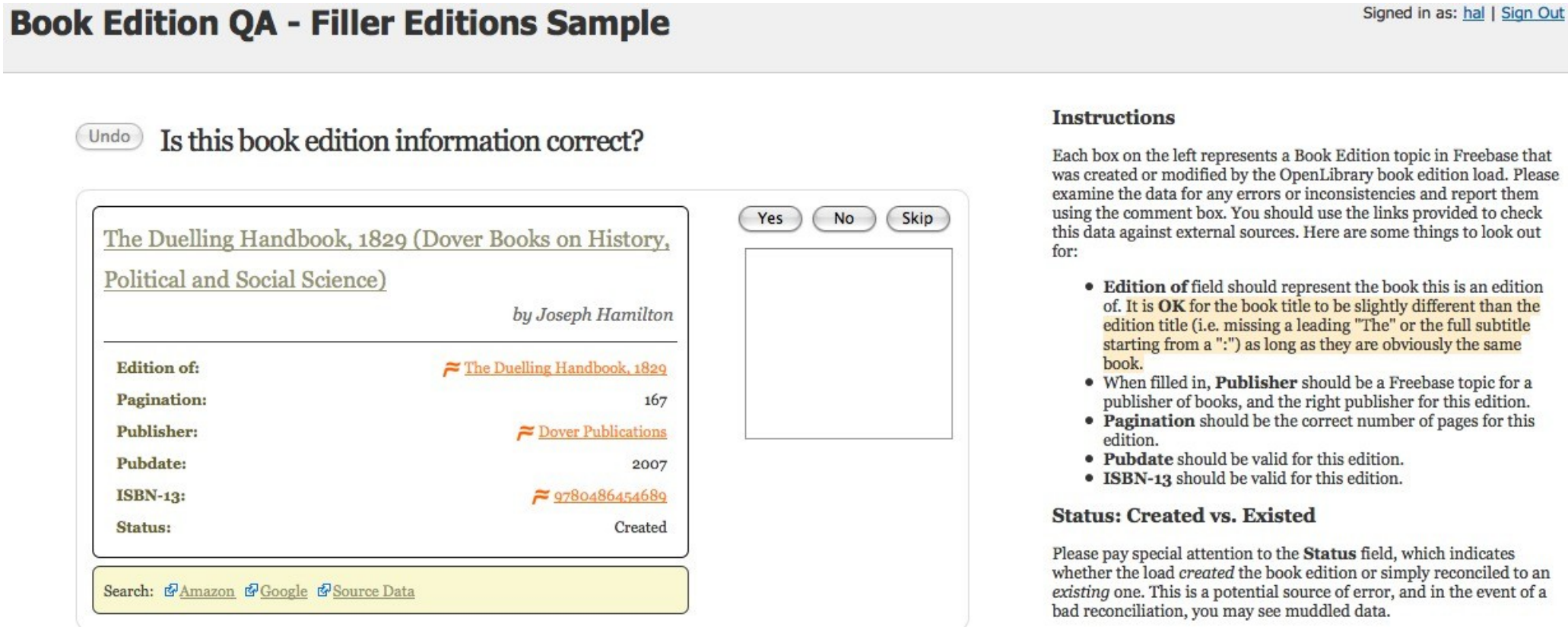
ALL TIME		1,142
1	mai_1980	229
2	mikan	165
3	jon	129
4	l_sangtam	122
5	eocay	110
6	kimgo	93
7	dericgayle	89
8	**kochhar**	**70**
9	anne8	66
10	jmontilla	56



Mass Data Loads

Precision: QA for >99% accuracy

Book Edition QA

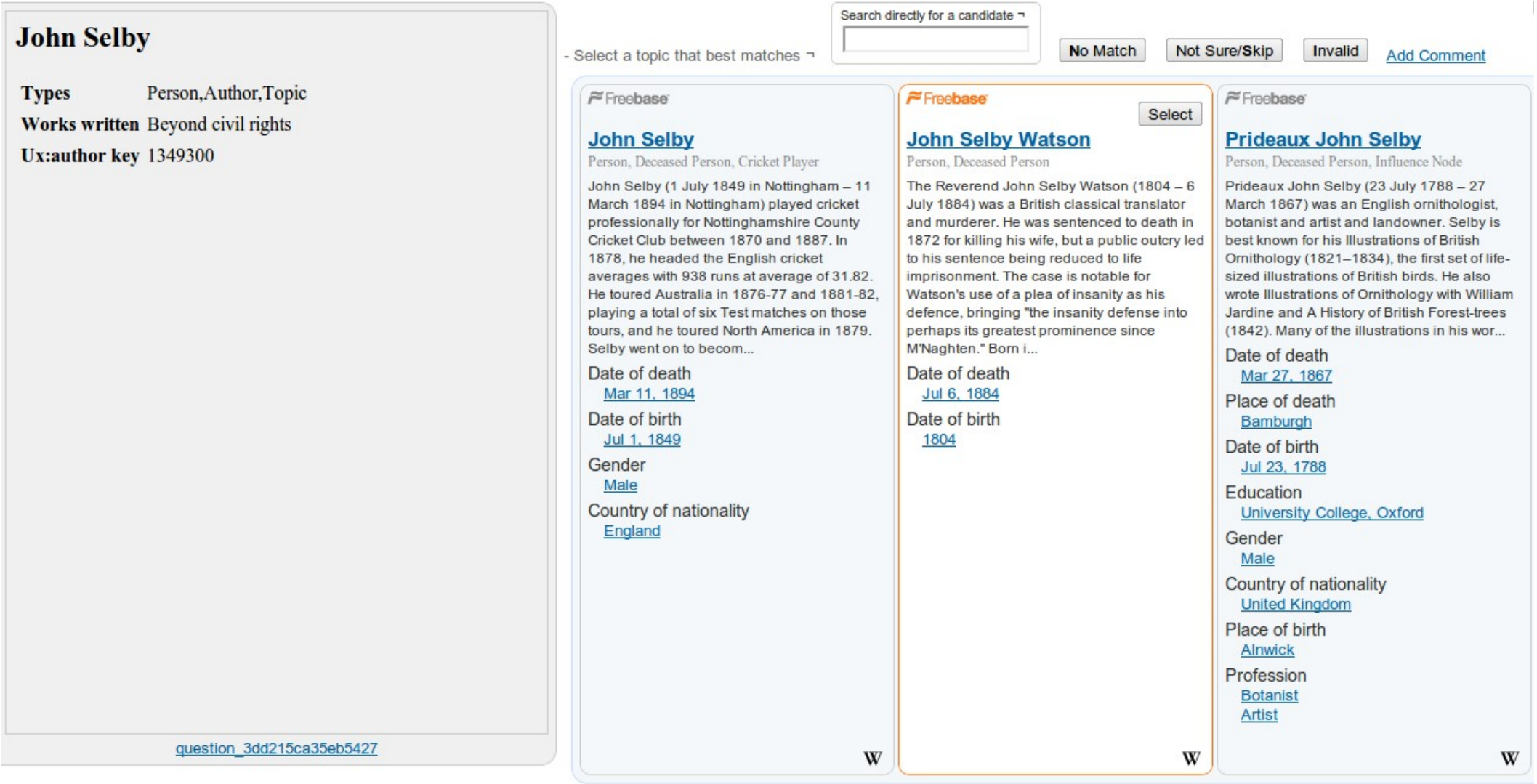


Mass Data Loads

Precision: QA for >99% accuracy

Coverage: Manual reconciliation

matchmaker



http://matchmaker2.freebaseapps.com/

1: Freebase & Human Computation

2: Example – Stanford Library

3: RABJ

4: Consensus

Reconcile Stanford Library Catalog with freebase.com

Aug 18, 2010

Freebase Meetup

Stanford Library Catalog

4.4MM book editions

1.3MM *English* book editions

1.2MM *English* books

600K authors

For freebase, identity is key

match books, match authors

Automatic matching insufficient

Trained judges needed to decide hard cases

How to get this?

RABJ

Redundant Array of Brains in a Jar

What?

Abstraction

Powers human judgment (HJ)

applications

3.1MM judgments in 16 months

Provides primitive elements for more

sophisticated applications

Questions

Judgments

Queues

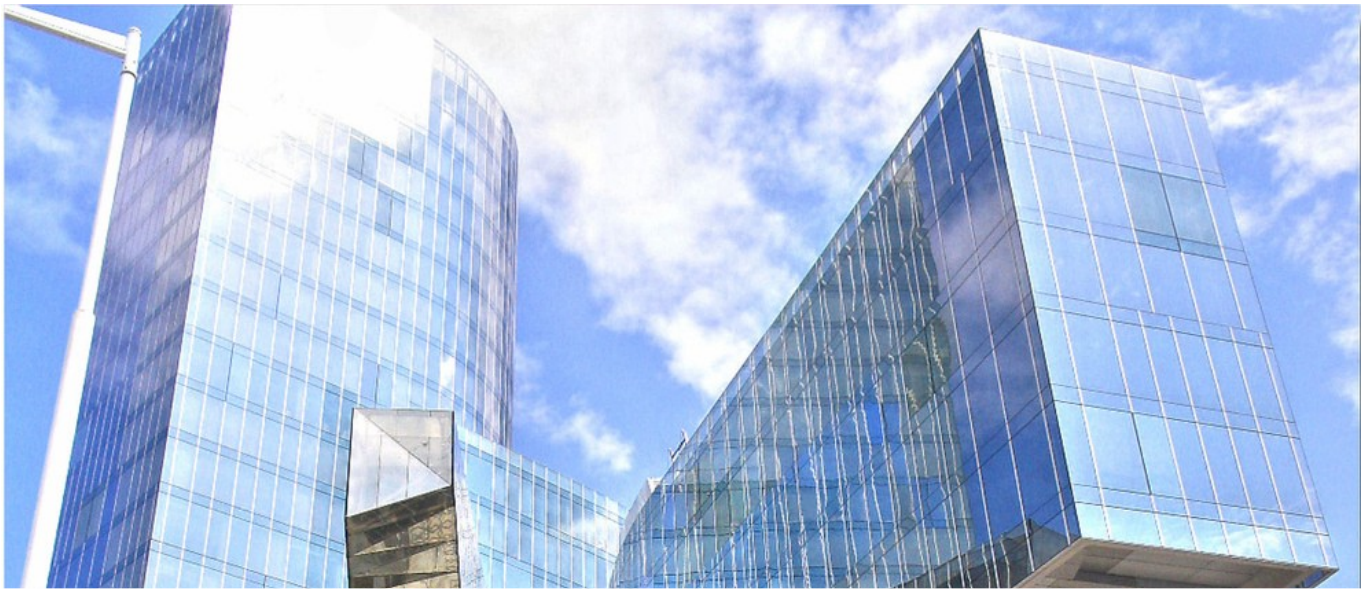
Agents

Design Constraints

Content-agnostic

Dynamic data

Low latency



Architecture

Questions contain metadata, pointers to dynamic content

Questions added to queues

Metadata allows slicing and dicing

Acre applications pull questions from RABJ

RABJ matches judge to available tasks

Acre renders question, sends judgment back

Declarative consensus
Yes: 3, No: 3, Skip: 4, Invalid: 3, Max: 6

RABJ notifies agents when consensus is reached

Scale

2.3 MM questions

3.1 MM judgments

500+ queues

20+ applications

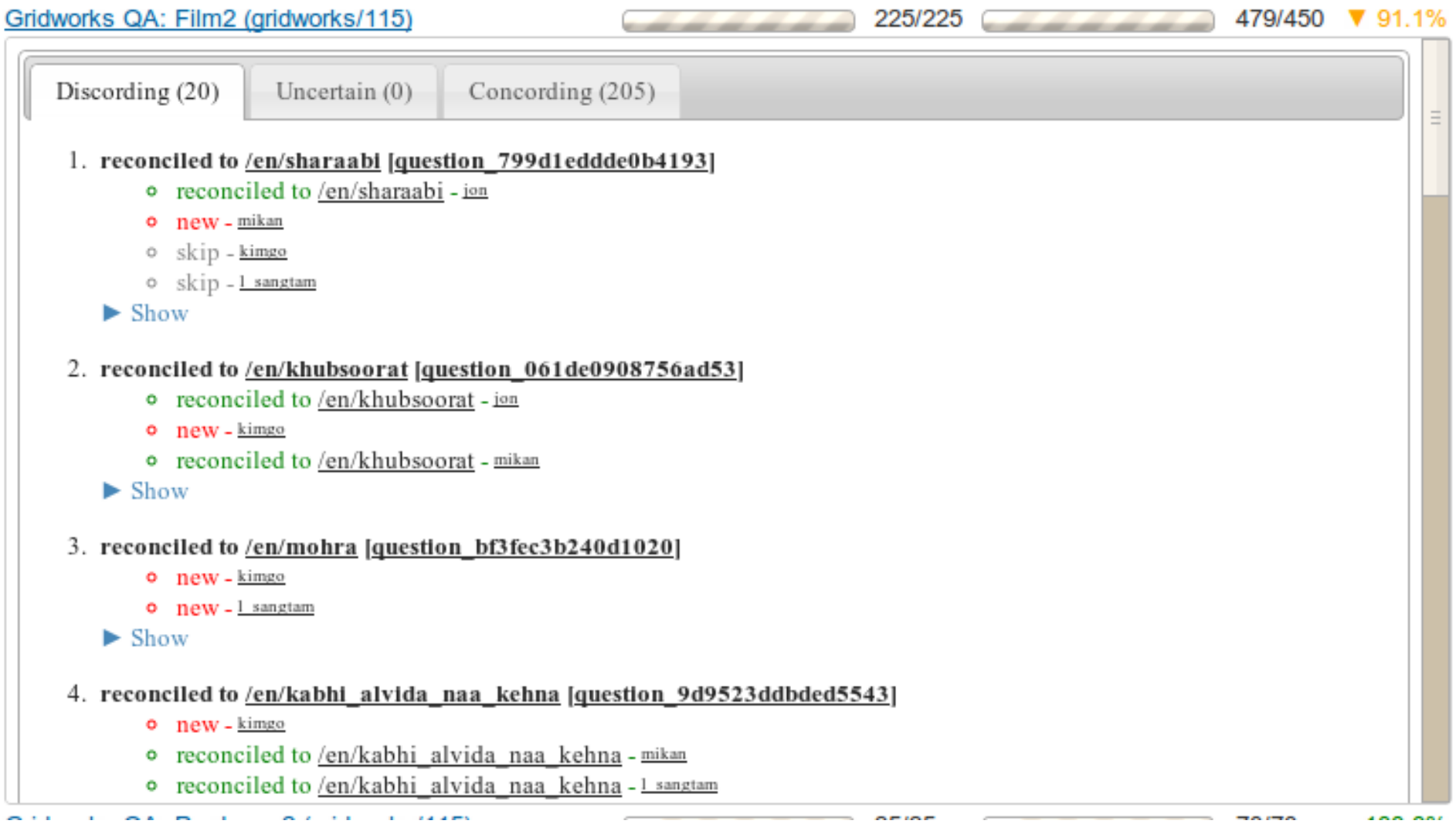
1: Freebase & Human Computation

2: Example – Stanford Library

3: RABJ

4: Consensus

Always have leftovers



Perfect Consensus? Not!

Evaluating QAers

Explore

http://rabj.freebaseapps.com/explorer

Create

http://wiki.freebase.com/wiki/RABJ_Tutorial

Reference

http://wiki.freebase.com/wiki/RABJ_API/

Questions?