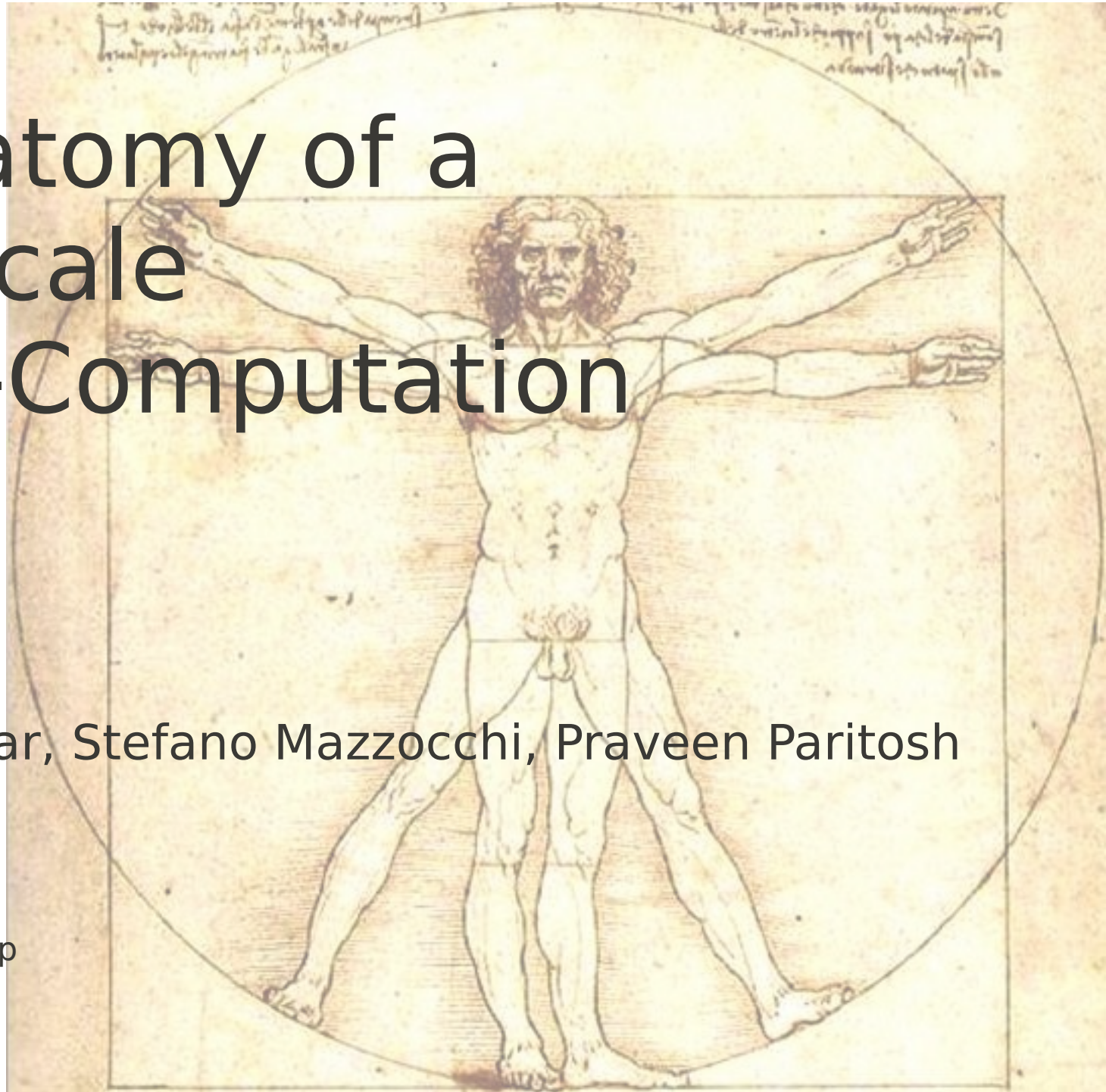


The Anatomy of a Large-Scale Human-Computation Engine

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?



Aug 18, 2010

Freebase Meetup



Community contributions

Mass Data Loads

Aug 18, 2010

Freebase Meetup



Human Judgments Improve Both



Aug 18, 2010

Freebase Mee



Community

Simplify contribution through games



TYPEWRITER



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*

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*

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

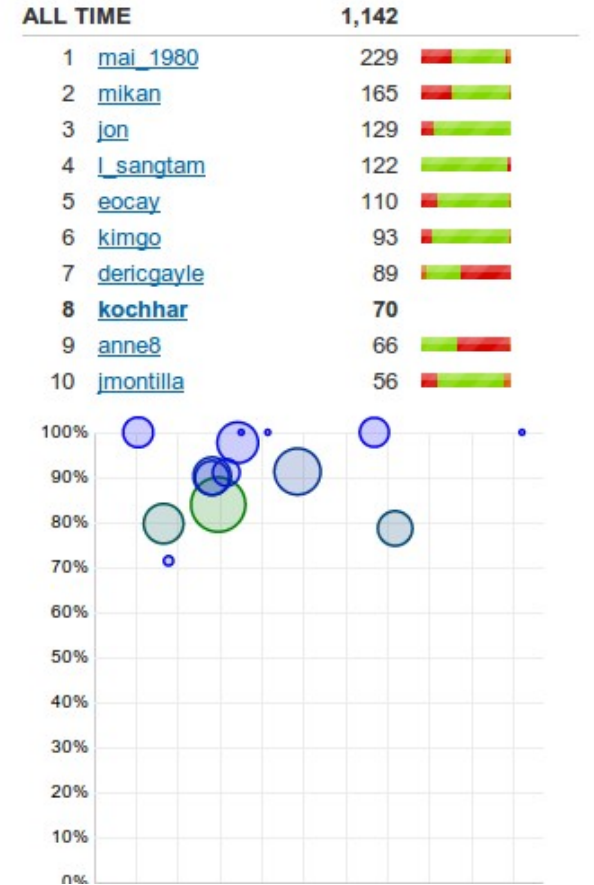


Started: a week ago
Finished: a week ago

The For Real Filmfare Best Movie Awards by [jon](#) ^{fb}

[▶ Show Load Data](#)

Queue	Answers	Judgements	Score
Gridworks QA: Producer 1 (gridworks/115)	155/155	332/310	93.5%
<p>Discarding (10) Uncertain (0) Concurring (145)</p> <p>1. reconciled to /en/mukesh_bhatt [question_6e3b5a8eddbc9703]</p> <ul style="list-style-type: none"> new - mikan skip - jmontilla reconciled to /en/mukesh_bhatt - mai_1980 reconciled to /en/mukesh_bhatt - eocav <p>▶ Show</p> <p>2. new [question_7dc89aa97fd5532b]</p> <ul style="list-style-type: none"> reconciled to /m/0clb6m6 - dfhuvnh reconciled to /m/0clb6m6 - eocav <p>▶ Show</p> <p>3. reconciled to /en/nazir_ahmed [question_6bd1dbf390d1d067]</p> <ul style="list-style-type: none"> new - eocav reconciled to /en/nazir_ahmed - mai_1980 skip - jmontilla skip - mikan <p>▶ Show</p> <p>4. reconciled to /en/mukesh_bhatt [question_0d4cca916effd364]</p> <ul style="list-style-type: none"> reconciled to /en/mahesh_bhatt - dfhuvnh 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%





Mass Data Loads

Precision: QA for $>99\%$ accuracy

Book Edition QA

Book Edition QA - Filler Editions Sample

Signed in as: [hal](#) | [Sign Out](#)

[Undo](#) Is this book edition information correct?

The Duelling Handbook, 1829 (Dover Books on History, Political and Social Science)

by Joseph Hamilton

Edition of:	The Duelling Handbook, 1829
Pagination:	167
Publisher:	Dover Publications
Pubdate:	2007
ISBN-13:	9780486454689
Status:	Created

Search: [Amazon](#) [Google](#) [Source Data](#)

[Yes](#) [No](#) [Skip](#)

Instructions

Each box on the left represents a Book Edition topic in Freebase that was created or modified by the OpenLibrary book edition load. Please examine the data for any errors or inconsistencies and report them using the comment box. You should use the links provided to check this data against external sources. Here are some things to look out for:

- **Edition of** field should represent the book this is an edition of. It is **OK** for the book title to be slightly different than the edition title (i.e. missing a leading "The" or the full subtitle starting from a ":") as long as they are obviously the same book.
- When filled in, **Publisher** should be a Freebase topic for a publisher of books, and the right publisher for this edition.
- **Pagination** should be the correct number of pages for this edition.
- **Pubdate** should be valid for this edition.
- **ISBN-13** should be valid for this edition.

Status: Created vs. Existed

Please pay special attention to the **Status** field, which indicates whether the load *created* the book edition or simply reconciled to an *existing* one. This is a potential source of error, and in the event of a bad reconciliation, you may see muddled data.



Mass Data Loads

Precision: QA for $>99\%$ accuracy

Coverage: Manual reconciliation



matchmaker

John Selby

Types Person, Author, Topic

Works written Beyond civil rights

Ux:author key 1349300

[question_3dd215ca35eb5427](#)

Search directly for a candidate ▾

- Select a topic that best matches ▾

No Match **Not Sure/Skip** **Invalid** [Add Comment](#)

Freebase

[John Selby](#)
Person, Deceased Person, Cricket Player

John Selby (1 July 1849 in Nottingham – 11 March 1894 in Nottingham) played cricket professionally for Nottinghamshire County Cricket Club between 1870 and 1887. In 1878, he headed the English cricket averages with 938 runs at average of 31.82. He toured Australia in 1876-77 and 1881-82, playing a total of six Test matches on those tours, and he toured North America in 1879. Selby went on to becom...

Date of death
[Mar 11, 1894](#)

Date of birth
[Jul 1, 1849](#)

Gender
[Male](#)

Country of nationality
[England](#)

W

Freebase

Select

[John Selby Watson](#)
Person, Deceased Person

The Reverend John Selby Watson (1804 – 6 July 1884) was a British classical translator and murderer. He was sentenced to death in 1872 for killing his wife, but a public outcry led to his sentence being reduced to life imprisonment. The case is notable for Watson's use of a plea of insanity as his defence, bringing "the insanity defense into perhaps its greatest prominence since M'Naghten." Born i...

Date of death
[Jul 6, 1884](#)

Date of birth
[1804](#)

W

Freebase

[Prideaux John Selby](#)
Person, Deceased Person, Influence Node

Prideaux John Selby (23 July 1788 – 27 March 1867) was an English ornithologist, botanist and artist and landowner. Selby is best known for his Illustrations of British Ornithology (1821–1834), the first set of life-sized illustrations of British birds. He also wrote Illustrations of Ornithology with William Jardine and A History of British Forest-trees (1842). Many of the illustrations in his wor...

Date of death
[Mar 27, 1867](#)

Place of death
[Bamburgh](#)

Date of birth
[Jul 23, 1788](#)

Education
[University College, Oxford](#)

Gender
[Male](#)

Country of nationality
[United Kingdom](#)

Place of birth
[Alnwick](#)

Profession
[Botanist](#)
[Artist](#)

W

<http://matchmaker2.freebaseapps.com/>

Aug 18, 2010

Freebase Meetup



1: Freebase & Human Computation

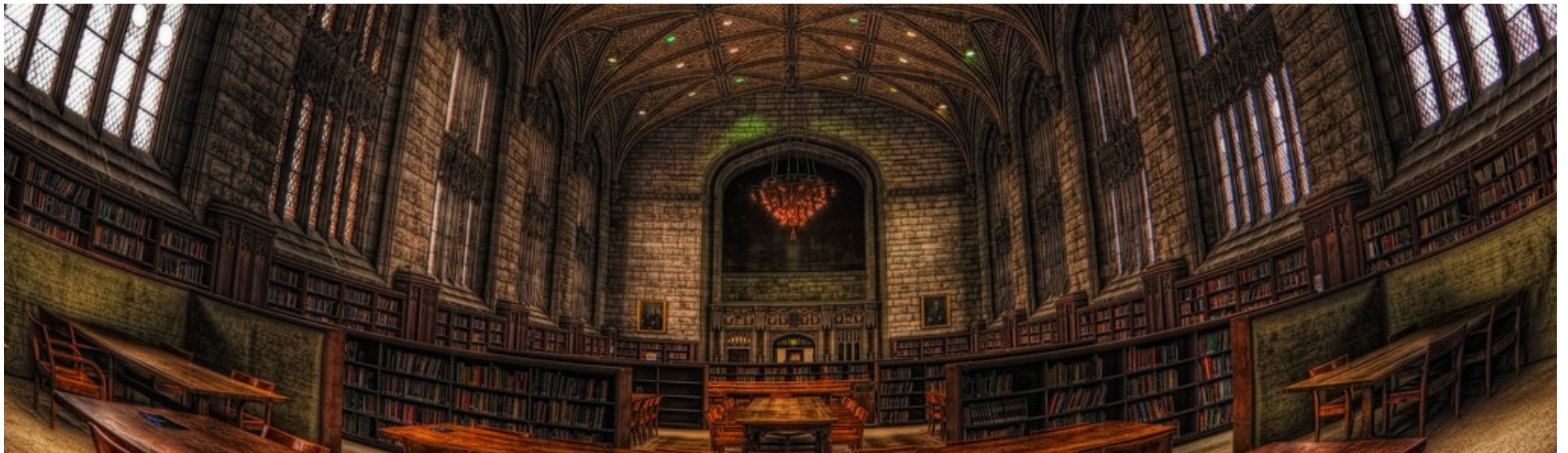
2: Example – Stanford Library

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?

Aug 18, 2010

Freebase Meetup



RABJ

Redundant Array of Brains in a Jar



What?

Abstraction

Powers human judgment (HJ)

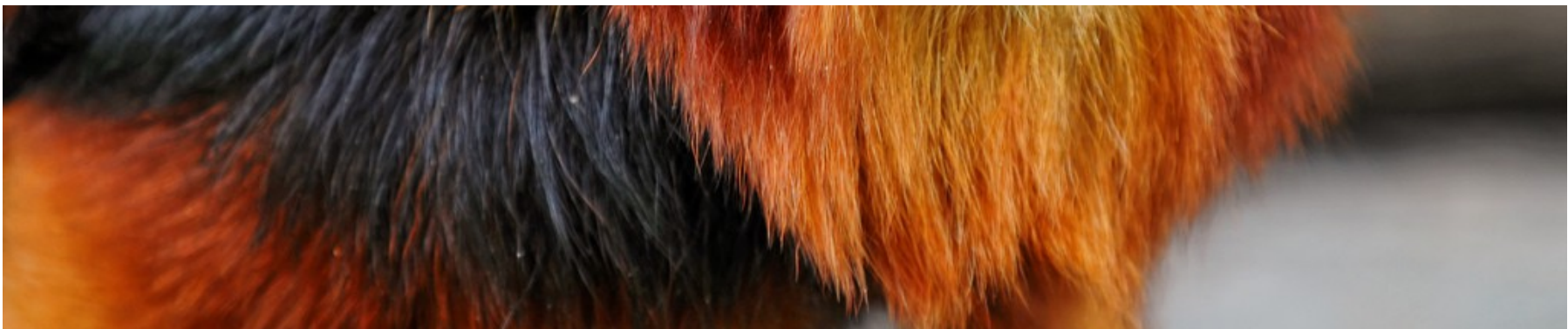
applications

3.1MM judgments in 16 months





Provides primitive elements for more sophisticated applications



Aug 18, 2010

Freebase Meetup





Questions Judgments Queues Agents

Aug 18, 2010

Freebase Meetup





Design Constraints



Aug 18, 2010

Freebase Meetup





Content-agnostic

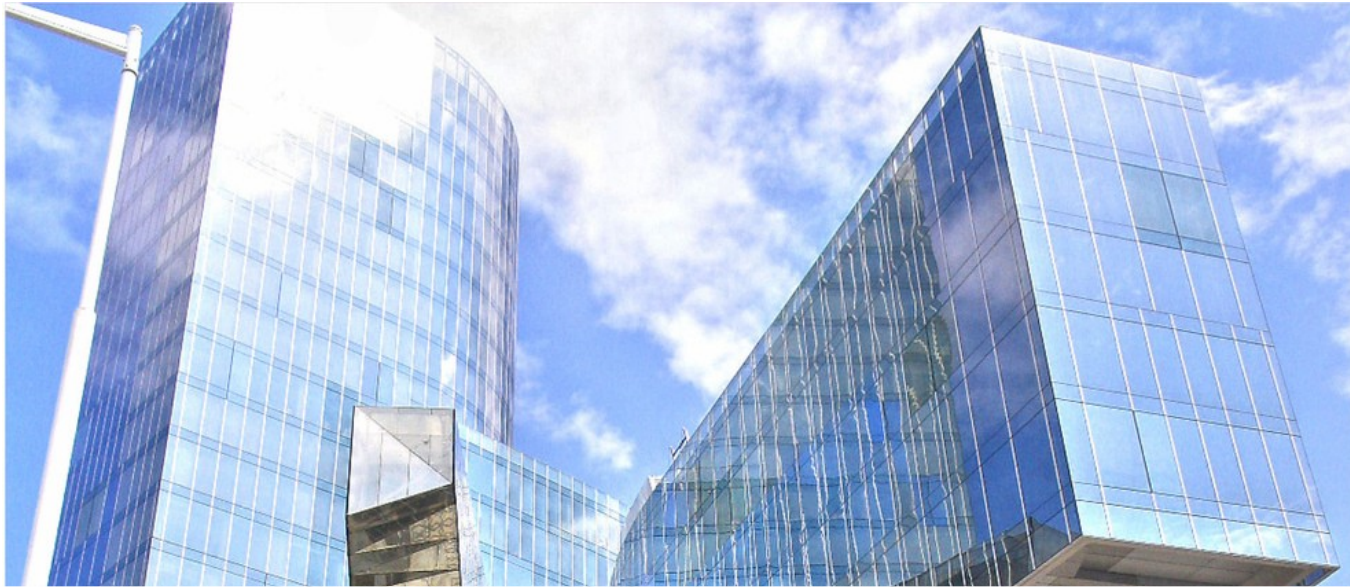
Dynamic data

Low latency

Aug 18, 2010

Freebase Meetup





Architecture



Aug 18, 2010

Freebase Meetup





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



Aug 18, 2010

Freebase Meetup





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

Aug 18, 2010

Freebase Meetup



Discarding (20)

Uncertain (0)

Concording (205)

1. reconciled to [/en/sharaabi](#) [question_799d1eddde0b4193]

- reconciled to [/en/sharaabi](#) - [ion](#)
- new - [mikan](#)
- skip - [kimgo](#)
- skip - [l_sangtam](#)

▶ Show

2. reconciled to [/en/khubsoorat](#) [question_061de0908756ad53]

- reconciled to [/en/khubsoorat](#) - [ion](#)
- new - [kimgo](#)
- reconciled to [/en/khubsoorat](#) - [mikan](#)

▶ Show

3. reconciled to [/en/mohra](#) [question_bf3fec3b240d1020]

- new - [kimgo](#)
- new - [l_sangtam](#)

▶ Show

4. reconciled to [/en/kabhi_alvida_naa_kehna](#) [question_9d9523ddbde5543]

- new - [kimgo](#)
- reconciled to [/en/kabhi_alvida_naa_kehna](#) - [mikan](#)
- reconciled to [/en/kabhi_alvida_naa_kehna](#) - [l_sangtam](#)

Perfect Consensus? Not!



Evaluating QAers

Aug 18, 2010

Freebase Meetup





Explore

<http://rabj.freebaseapps.com/explorer>

Create

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

Reference

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





Questions?