DATA-CENTERED CROWDSOURCING WORKSHOP

PROF. TOVA MILO

SLAVA NOVGORODOV

TEL AVIV UNIVERSITY 2014/2015

PROJECTS

- The team size is around 4 students and should be decided by the next week
- The projects should be done during the semester
- There is one “touch-base” meeting in a month and submission at the last week of the semester (demo day)
- The topic and the technology is up to you (we teach web programming using PHP, JavaScript and HTML and provide an SDK)
- The project should be interesting and solve real problem in the world of crowdsouring

PROJECTS

- In general there are two possible options for the projects:
  - Application (Web or Mobile) – an app that solves concrete problem using crowdsourcing (for example: Application for text translation)
  - Infrastructure – provides an option for application to integrate with it, and solves more general crowdsourcing problem (for example: Platform for matching users to questions/tasks)
- In both cases, students should prepare an end-to-end demonstration of their projects

CROWDSOURCING APPLICATION

- Possible options for applications projects:
  - Tagging images (or in general: Adding metadata to multimedia content) using crowd
  - Text translation, adaptation, shortening, improvement (TL;DR)
  - Answering questions/queries using crowd
  - Planning and scheduling with the crowd
  - Geo-spartial applications based on crowd (use crowd locations to improve your data)
  - Building knowledge base (incl. cleaning)
  - …

CROWDSOURCING INFRASTRUCTURE

- Possible options for infrastructure projects:
  - Crowdsourcing platform design and implementation (AMTurk, CrowdFlower, …)
  - Matching crowd users to questions/tasks (based on their preferences/test questions/performance)
  - Payment and incentives model for crowd (solving the problem of trade-off between budget and quality)
  - Budget management and payment strategies
  - Ordering/Sequences of questions to crowd (session model and context)
  - Spamming detection
  - Load balancing

COLLECTING META-DATA FROM MULTIMEDIA

- Existing projects:
  - The ESP game – tagging images
  - Tag a Tune – audio annotation
  - Squigl – detect objects on images
  - Verbosity – a common knowledge trainer (game)
  - More projects – Games with a Purpose
- More ideas for projects:
  - Video tagging, events detection using crowd

THE ESP GAME

score
6100
ESP Game
Concentrate...
time
0:14
What do you see?
taboo words
forest
meadow
guesses
tree
grass
Matched on: tree
submit
pass

TAG A TUNE

Score
50
Bonus
Tag a Tune
Hear Here
Timer
1:48
Describe the tune ...
0:11
Listening to the same tune?
same
different
1 in a row!
your descriptions
alarm
your partner's descriptions
wierd
wierd wind instrument
alien
submit
pass

SQUIGL

score
586
Squigl
Ready, set... Trace!
time
0:25
Trace the...
eye
Partner Submitted!
pass
submit
auto-submit when done

BONUS!
5,000 PTS

Squigl
1175
1:26
Trace the...
clothes
auto-submit when done

VERBOSITY

gwap
ESP Game
Tag a Tune
Verbosity
Squigl
Matchin
Refer a friend for 15,000 pts!
elianealhadef
Verbosity
it's common sense.
How to Play
1 You and a partner alternate between being Describer and Guesser.
2 As Describer, you must help your partner guess the secret word by giving clues.
it contains a
3 As Guesser you must type the secret word that your partner is describing.
new guess
movie
Got it, Let's Play!
View Top Scores

TEXT TRANSLATION

- Existing projects:
  - Paper from UPenn by Zaidan & Callison-Burch
  - Facebook localization was partially done by crowd
  - GetLocalization.com
  - TL;DR community
  - Google translate “improve translation” feature
- More ideas for projects:
  - Projects like English Wikipedia → Simple English Wiki

GET LOCALIZATION

Crowdsourcing the Right Way

- Give the translators enough information about the translatable materials. This is especially important for new features.
- Tell them what the new features do and to which features specific strings belong. Anticipate difficult text parts and give detailed instructions about them.
- You can add information about the strings in the editor's Context field. Do this directly in the editor or by using comment field of the resource files.

- Add a link to the localization project in your app, tell about it on your webpage and discussion forums, and find users in social media (Twitter, Facebook, Weibo, Vkontakte etc.)
- Encourage your users to brag about their translations by using the dedicated buttons in Get Localization.
- Go fully open and invite everybody to translate or restrict who can translate for you: make your project Public or Private. In both alternatives you can give your trusted users more rights and assign them e.g. as Moderators.

- Answer questions, moderate discussions. Never let your community wait for too long.
- You get translators' questions directly to your email and can move straight to the string in question. If you reply quickly, the translators know the project is important for you!

- Often a simple 'Thank you' might be the best reward, just don't forget to say it! Other possibilities include badges or beta user rights, or mentioning people in your app's credits.
- In the project statistics you can see your translators and their translated amounts language by language.

TL;DR FACEBOOK COMMUNITY

אמ;לק поделился(лась) фотография רשת.
13 Март в 22:46 ·

לא ידוע.

רשת

לאן נעלם ולדימיר פוטין?

http://s.reshet.tv/18iyar5

אמ;לק поделился(лась) фотография מעריב אונליין
Maariv Online.
16 Март в 20:07 ·

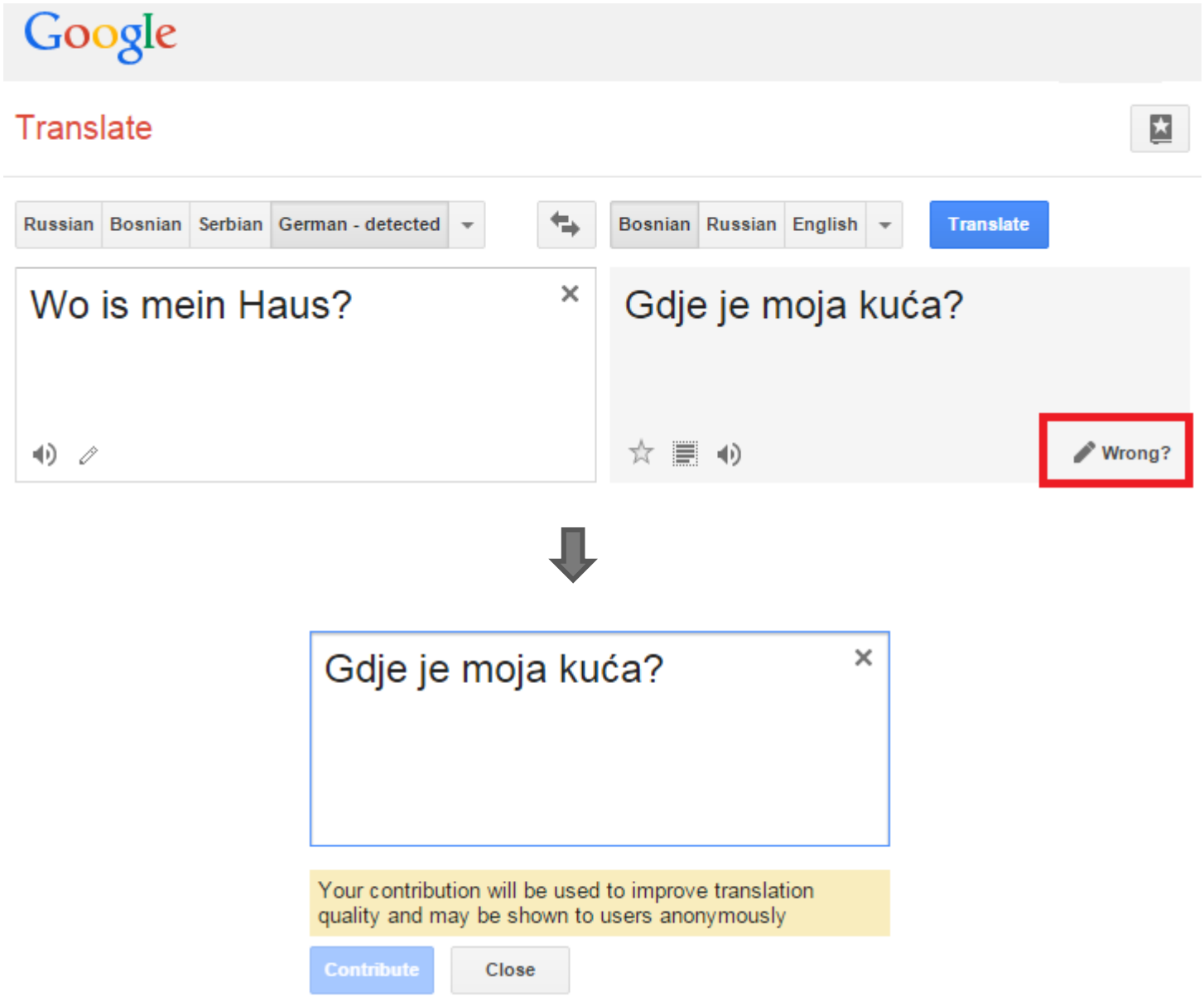
משכנעים סטודנטים ערבים להצביע להם, ומארגנים מערך הסעות. גם לסטודנטים הלומדים בירדן.
האמת? בניגוד להבטחה, אנחנו מאמינים.

מעריב אונליין Maariv Online

לא תאמינו מה הרשימה המשותפת עושה כדי להביא קולות

http://www.maariv.co.il/news/elections-2015/Article-468360

GOOGLE TRANSLATE



ANSWERING QUESTIONS AND QUERIES WITH CROWD

- Existing projects:
  - StackOverflow
  - TripAdvisor
  - Yahoo! Answers
  - A lot of “academical” research to make such projects more structured and focused (OASSIS, AskIt, …)
- More ideas for projects:
  - Free text to queries translation (using crowd)
  - Meta-queries

PLANNING WITH CROWD

- Existing projects:
  - CrowdPlannr (academical research, TAU)
  - CrowdCierge (also academical, MIT)
- More ideas for projects:
  - Events planning
    - Schedule
    - Suggestions
    - Work split and sharing
  - Better (more general) CrowdPlannr…

CROWDPLANR

PLANNING MADE EASY WITH CROWD

BUILD ME A TRIP | **CONTRIBUTE** | STATISTICS | SEARCH TRIPS | HELP

Welcome, **Dennis**
(2 points)

Region: **Central Europe**
Period: **September**
Duration: **1 month**
Traveller: **2 friends**
Level: **cities**
Request: **We hired a car and want to do a series of short (one day) hikes in the Alpine scenery. We land and fly out from Milan, Italy.**

Current path: **Canazei - Meyerhofen - Konigsee - ...**

Please advise the next step:

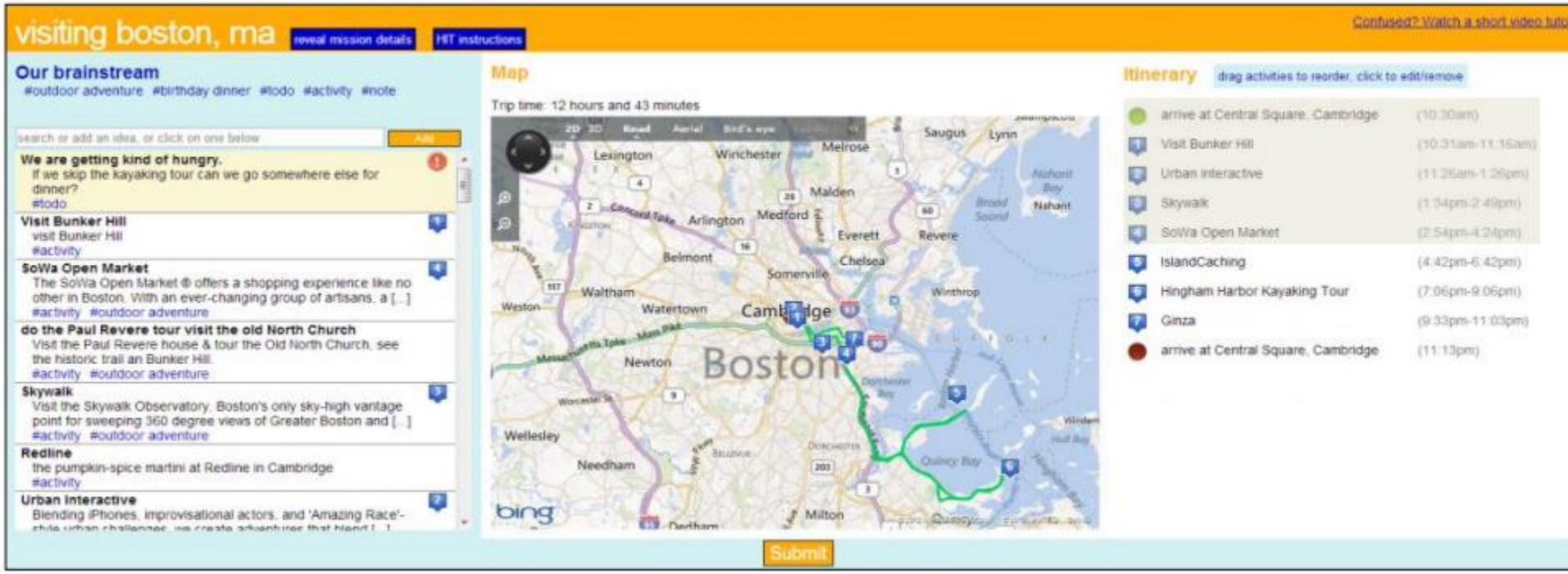
I

Interlaken, Switzerland
[Stop at Interlaken]

Innsbruck, Austria
[Stop at Innsbruck]

I - Add new!

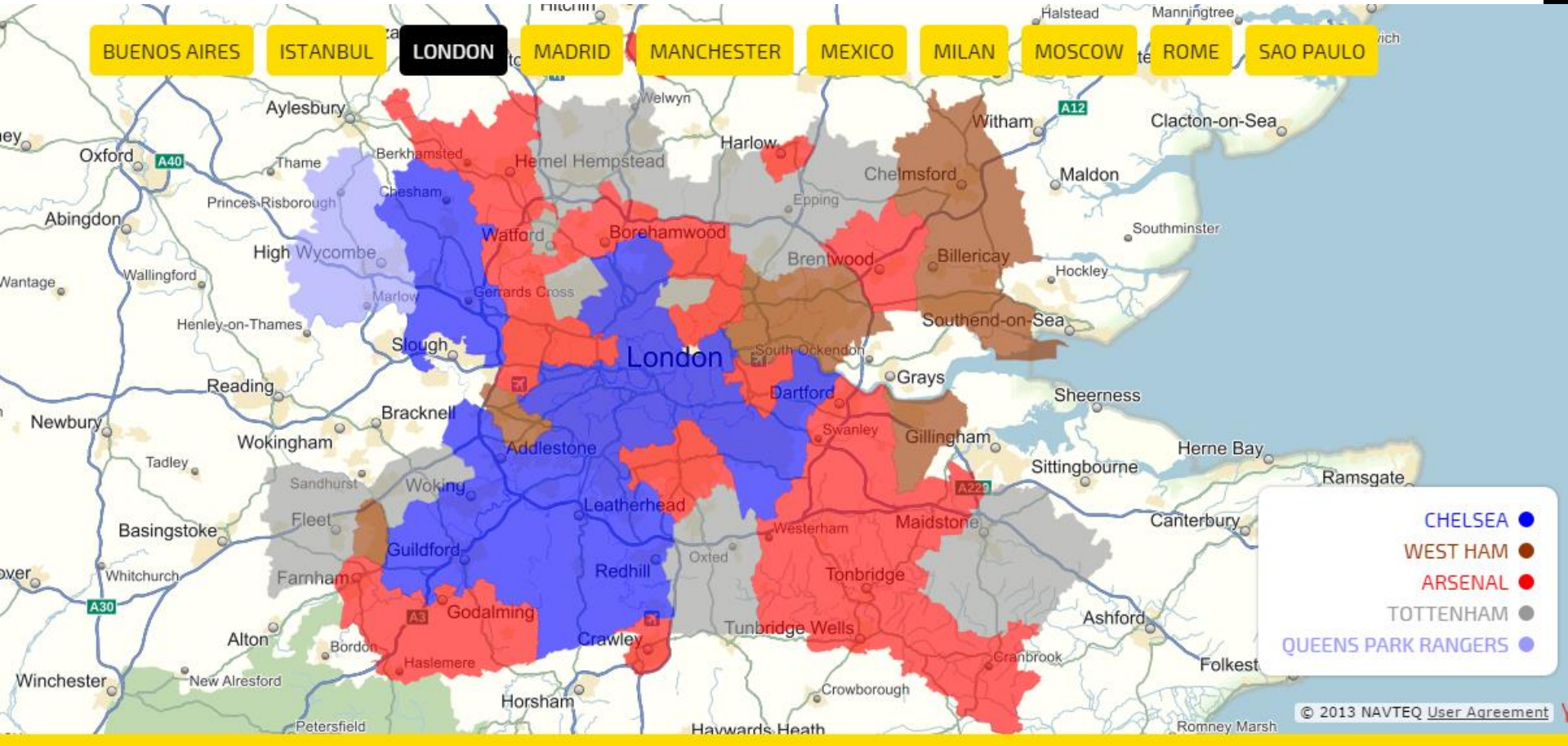
CROWDCIERGE



GEOSPATRIAL CROWD APPS

- Existing projects:
  - Scores & Video city maps – soccer fans visualization
  - Waze
  - OpenStreetMaps – collaborative mapping
- More ideas for projects:
  - Wikipedia missing photos
    - If there is a missing photo of some place, you can ask user to upload it (if he is currently there)
  - Geo-based tasks and job sequence planning
  - Crowd-based layers over existing maps

SCORES&VIDEO CITY MAPS



SCORES&VIDEO

SOURCE: FOOTBALL FANS' INSTAGRAMS
SEE ALL OF THEM LIVE IN **SCORES&VIDEO FOR IOS**

WANT S&V FOR ANDROID?

18 0

WAZE

9:57
64%
30 מ'
הרצל
נהריה
הנרייטה סולד
שופרסל
הרצל
צבעוני
נופר
אחד העם
שטיינהרט
שדרות קיי
בית הבראה
בית קיי
ש"י עגנון
ברקן
רותם
הסייפן
0
10:30
31 ק"מ
32 דק'
28
קמ"ש

OPENSTREETMAPS

TZEVAADOM.COM

BUILDING KNOWLEDGEBASE

- Existing projects:
  - YAGO (academical, not crowdsourcing)
  - OASSIS (academical, crowd minning)
  - QOCO (academical, query-oriented)
- More ideas for projects:
  - Building database of popular facts (common knowledge)
  - Building database of “rare”/not documented facts
  - Enrich datasets with semantics
  - Combine multipledatasets

OASSIS

(1) - **User panel**. Used for logout and sh
A bit more about rating:
= 5 answers
=
=

(2) - **Never go/eat there**. Used to specify
not relevant for you and the system should
question about this place/restaurant

(3) **Submit**. First, you should select the f
current pair and you can click on this butto

OR

(4) **Specify**. Used if you want to give a sp
example. The frequency of more general q
submitted.

Got the rules, let me play!

ASKIT

AskIt!
Asking the Right Quetions
AskIt!
Admin View
Statistics
Logout
(5/10)
Hosagrahar V Jagadish
University of Michigan
DBLP
Who does Hosagrahar V Jagadish resemble most
Kevin Spacey
David Beckham
Barack Obama
Is Hosagrahar V Jagadish theoretician or practitioner?
Theoretician
Practitioner
Both
Is Hosagrahar V Jagadish senior or junior researcher?
Senior
Junior
In between

TEASERS

01 home 02 learned rules 03 user answers 04 answer questions 05 about

macademo
crowdsourcing well-being

search

search

rules of the day

morning -> jogging

the flu -> chicken soup

academia

Full Paper

TAU DB Group

The MoDaS Project

the flu -> chicken soup

Correlation

Items correlated with *the flu* :

garlic

Items correlated with *chicken soup* :

a cold

the flu

You might find these interesting...

Key Facts About Flu and Flu Vaccine

Scientific American – Influenza

25 Comforting Chicken Soup Recipes

Chicken Soup Scientifically Proven to Help Cure Colds

GUESS & TRIVIA

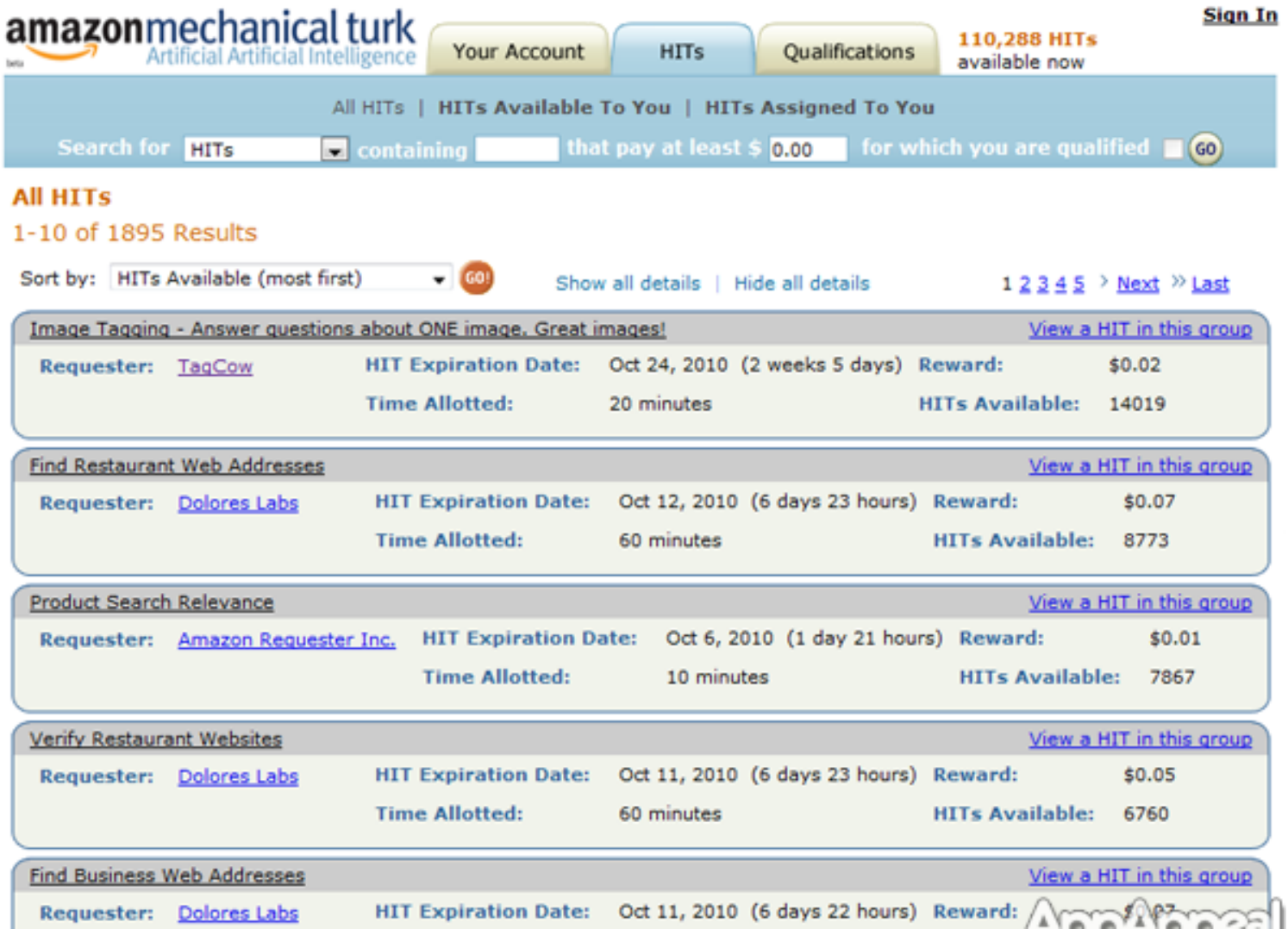
CROWDSOURCING INFRASTRUCTURE

- Possible options for infrastructure projects:
  - Crowdsourcing platform design and implementation (AMTurk, CrowdFlower, …)
  - Matching crowd users to questions/tasks (based on their preferences/test questions/performance)
  - Payment and incentives model for crowd (solving the problem of trade-off between budget and quality)
  - Budget management and payment strategies
  - Ordering/Sequences of questions to crowd (session model and context)
  - Spamming detection
  - Load balancing

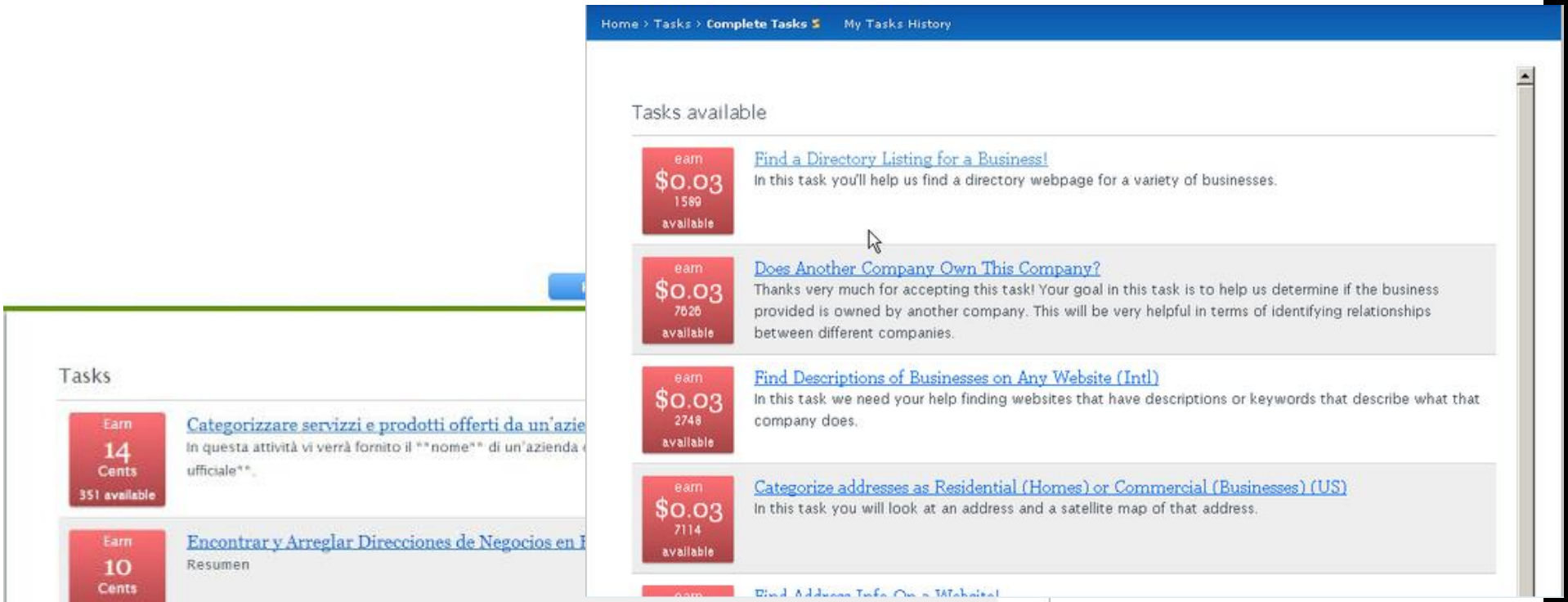
CROWDSOURCING PLATFORM

- Existing projects:
  - Amazon Mechanical Turk (microtasks)
  - CrowdFlower
  - oDesk (for big projects)
- More ideas for projects:
  - Better management panel
  - Better connection to users

MECHANICAL TURK



CROWDFLOWER



MATCHING CROWD USERS TO QUESTIONS/TASKS

- The matching can be done based on:
  - User preference
  - Requester preference (“Give me 5 people from US, that are experts in baseball”)
  - Qualification tests
  - More?

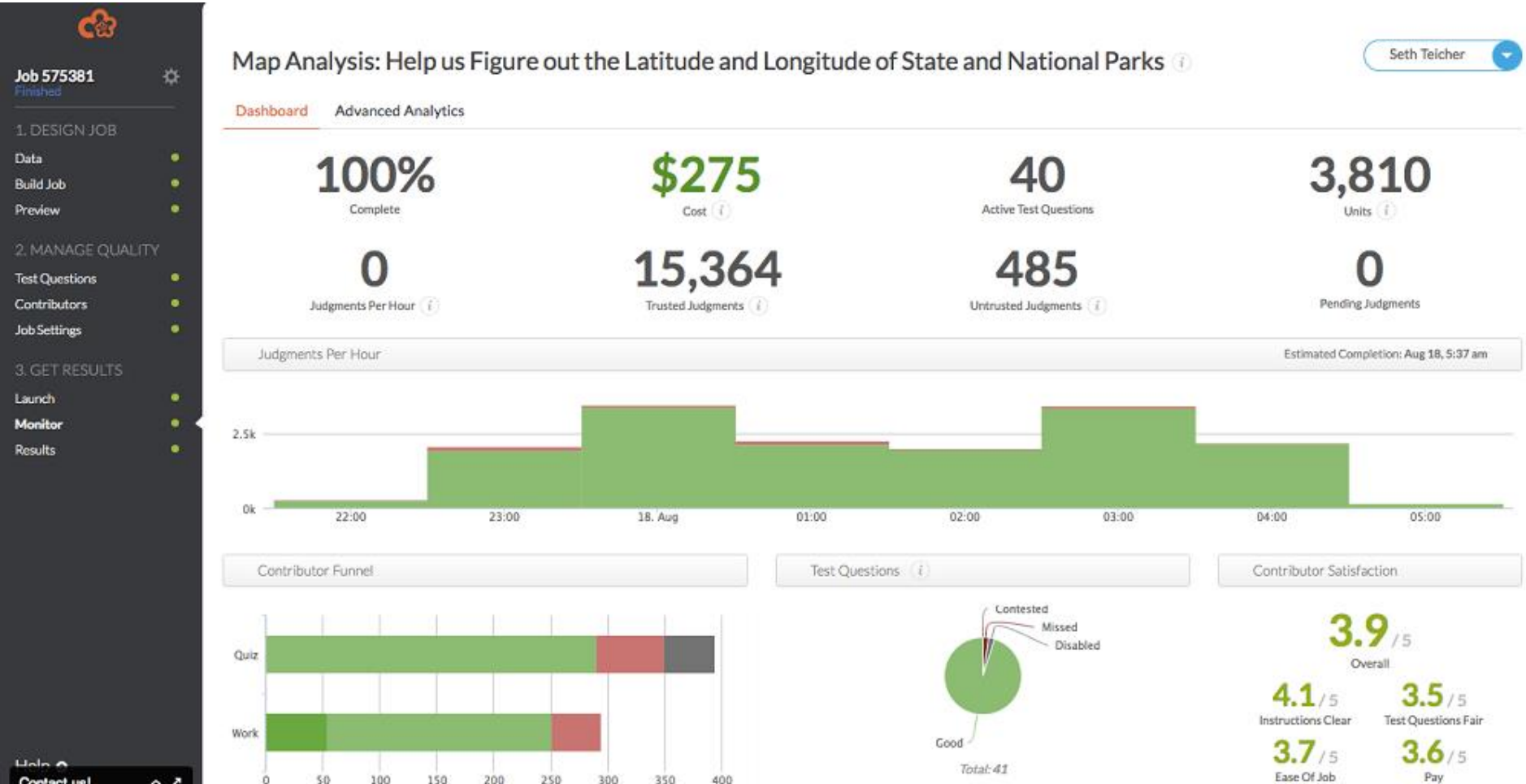
PAYMENT AND INCENTIVES

- Create the real “free” market:
  - Bidding
  - Monopoles and workers unions
  - Long-term contracts
  - Quality / Payment tradeoff
  - …

BUDGET MANAGEMENT

- Easier way to manage your workers
  - Algorithmic budget allocation
  - Manual preferences
  - Better UI (control panel)
  - …

MANAGEMENT PANEL



ORDERING OF TASKS

- Crowd users have different preferences
  - Many tasks of the same type
  - Many tasks in the same context
  - Different tasks
  - “Easy-Only” (less money per tasks – more tasks)
  - “Difficult-Only”
  - Mixed
  - …

SPAMMING DETECTION

- Detect spammers
- Don’t hurt real crowd (can cost you reputation)
- Provide high quality and low latency
- …

LOAD BALANCING

- The tasks should be done by all the crowd…
- There are many “strong” crowd users, and many requesters willing to get them
- Not always available
- …

Questions?