

DATA-CENTERED CROWDSOURCING WORKSHOP

PROF. TOVA MILO

SLAVA NOVGORODOV

TEL AVIV UNIVERSITY 2016/2017

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 crowdsourcing

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-spatial 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**

Timer **1:48**

Bonus 

Tag a Tune

Hear Here

Describe the tune ...

 0:11   

your descriptions

alarm

your partner's descriptions

wierd

wierd wind instrument

alien

Listening to the same tune?

SQUIGL

score
586

 Squigl
Ready, set... Trace!

time
0:25

Trace the...



auto-submit when done



BONUS!
5,000 PTS ▶

score
1175

 Squigl
Ready, set... Trace!

time
1:26

Trace the...



auto-submit when done



VERBOSITY

gwap

ESP Game

Tag a Tune

Verbosity

Squigl

Matchin

Refer a friend for 15,000 pts!

elianealhade f



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



Prepare Well

- 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.

Get Publicity & Build a Team

- 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.

Be Active & Engaged

- 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!

Don't Forget Rewards

- 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



Amikam поделился(лась) фотографія רשת.

13 Март в 22:46 · 🌐

לא ידוע.



רשת

לאן נעלם ולדימיר פוטין?

<http://s.resnet.tv/18iyar5>



מעריב אונליין поделился(лась) фотографія

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



Translate



Russian Bosnian Serbian German - detected



Bosnian Russian English

Translate

Wo is mein Haus?



Gdje je moja kuća?



Wrong?



Gdje je moja kuća?



Your contribution will be used to improve translation quality and may be shown to users anonymously

Contribute

Close

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:

 **Interlaken, Switzerland**
[Stop at Interlaken]

 **Innsbruck, Austria**
[Stop at Innsbruck]

 **I - Add new!**

CROWDCIERGE

Confused? Watch a short video tuto

visiting boston, ma

review mission details

HT instructions

Our brainstream

#outdoor adventure #birthday dinner #todo #activity #note

search or add an idea, or click on one below

Add

We are getting kind of hungry.

If we skip the kayaking tour can we go somewhere else for dinner?
#todo

Visit Bunker Hill

visit Bunker Hill
#activity

SoWa Open Market

The SoWa Open Market ® offers a shopping experience like no other in Boston. With an ever-changing group of artisans, a [...]
#activity #outdoor adventure

do the Paul Revere tour visit the old North Church

Visit the Paul Revere house & tour the Old North Church, see the historic trail an Bunker Hill
#activity #outdoor adventure

Skywalk

Visit the Skywalk Observatory, Boston's only sky-high vantage point for sweeping 360 degree views of Greater Boston and [...]
#activity #outdoor adventure

Redline

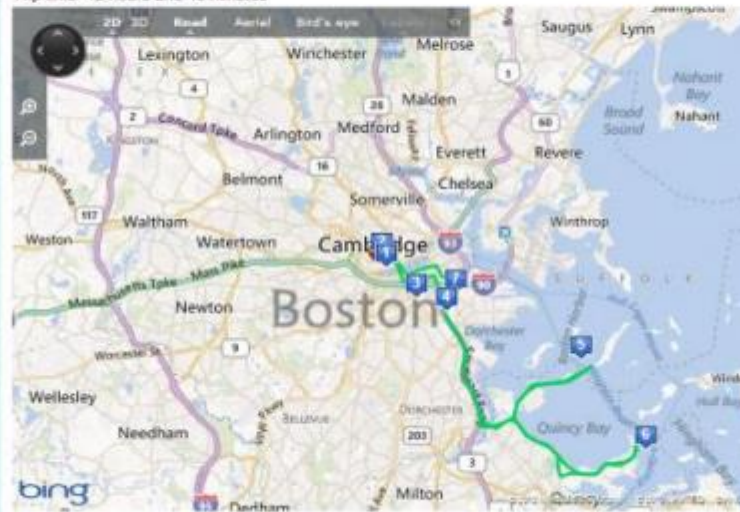
the pumpkin-spice martini at Redline in Cambridge
#activity

Urban Interactive

Blending iPhones, improvisational actors, and 'Amazing Race'-style urban challenges, we create adventures that blend [...]

Map

Trip time: 12 hours and 43 minutes



Submit

Itinerary

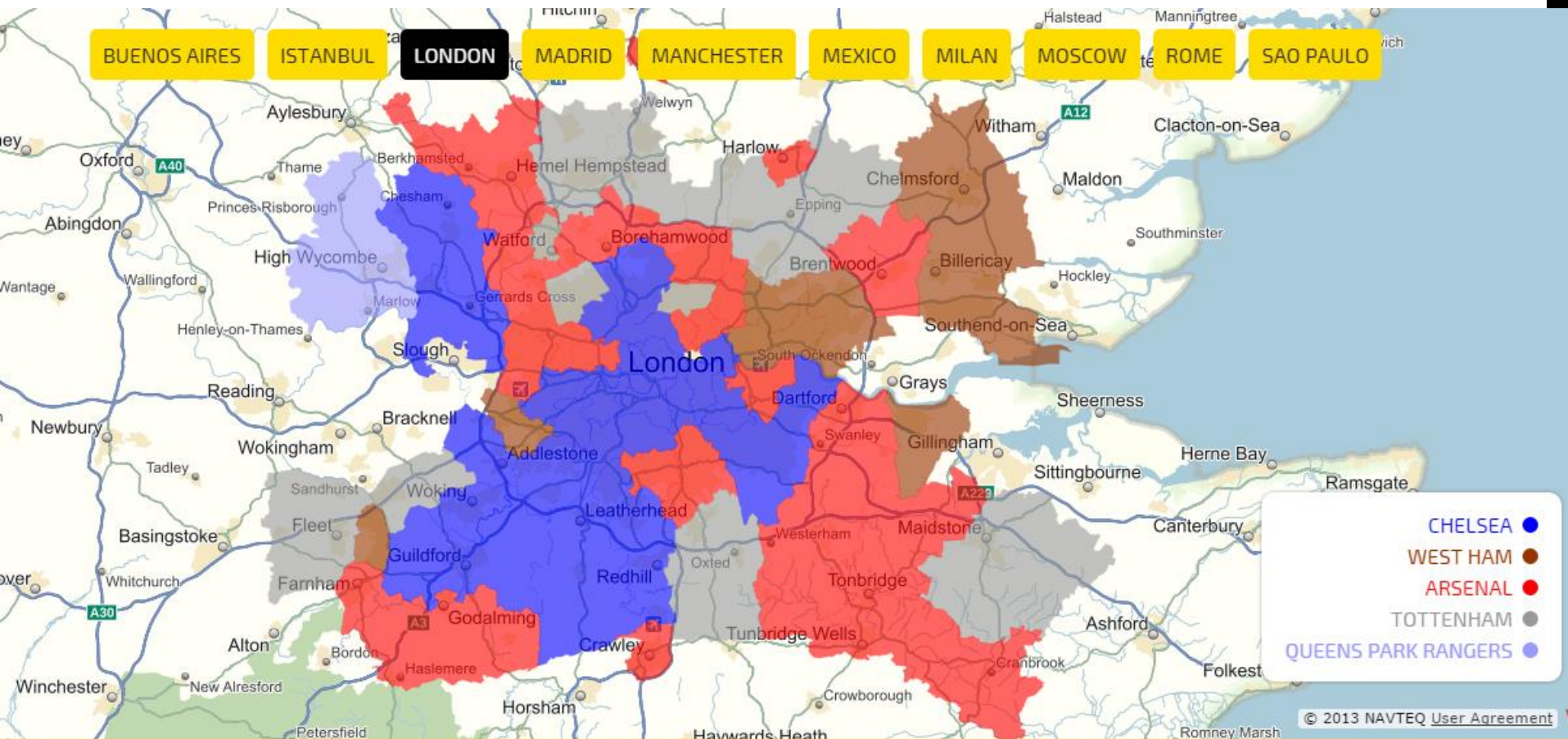
drag activities to reorder, click to edit/remove

- arrive at Central Square, Cambridge (10:30am)
- 1 Visit Bunker Hill (10:31am-11:15am)
- 2 Urban interactive (11:25am-1:25pm)
- 3 Skywalk (1:34pm-2:49pm)
- 4 SoWa Open Market (2:54pm-4:24pm)
- 5 IslandCaching (4:42pm-6:42pm)
- 6 Hingham Harbor Kayaking Tour (7:06pm-9:06pm)
- 7 Ginza (9:33pm-11:03pm)
- arrive at Central Square, Cambridge (11:13pm)

GEOSPATIAL 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

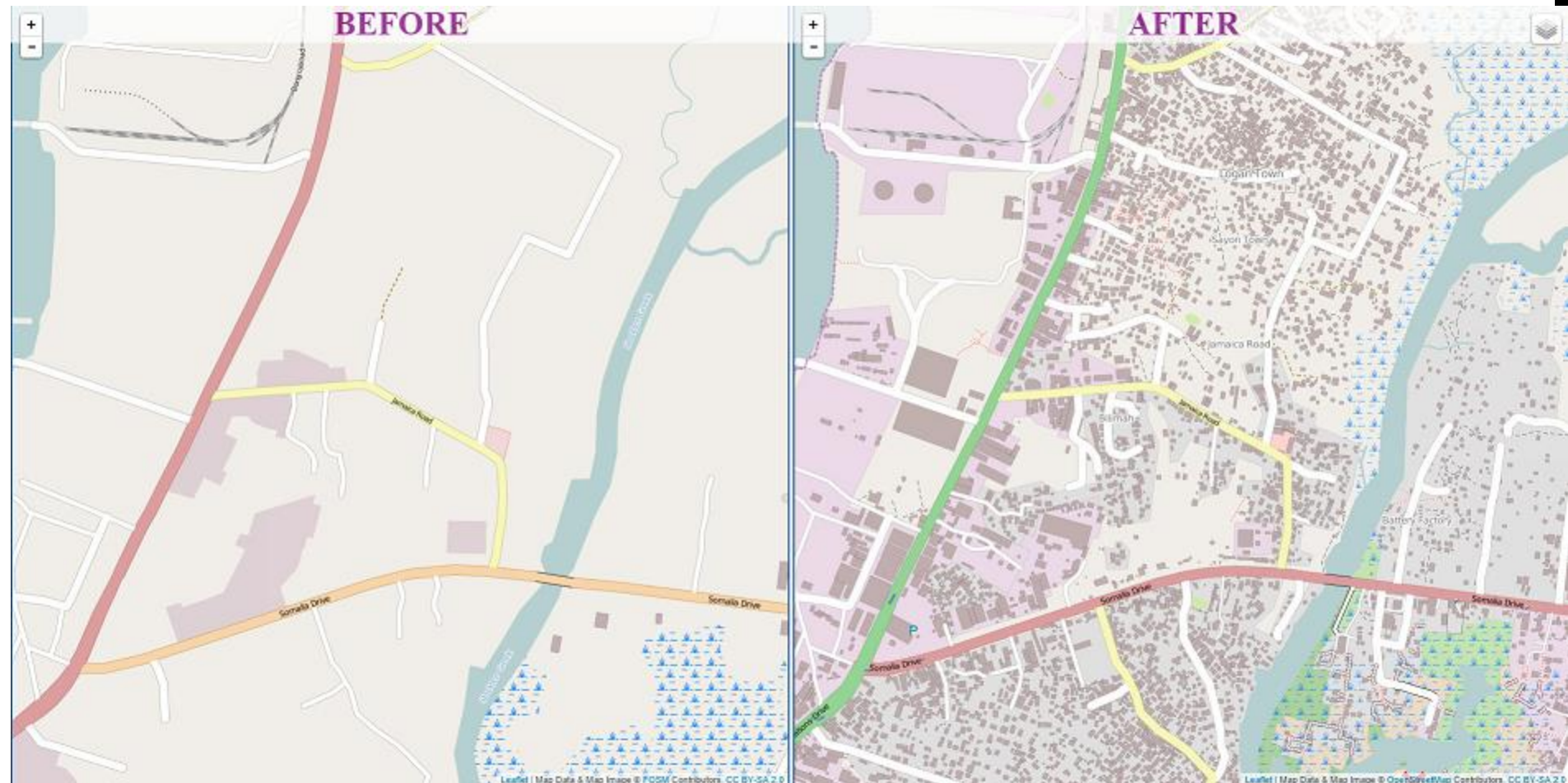
SCORES & VIDEO CITY MAPS



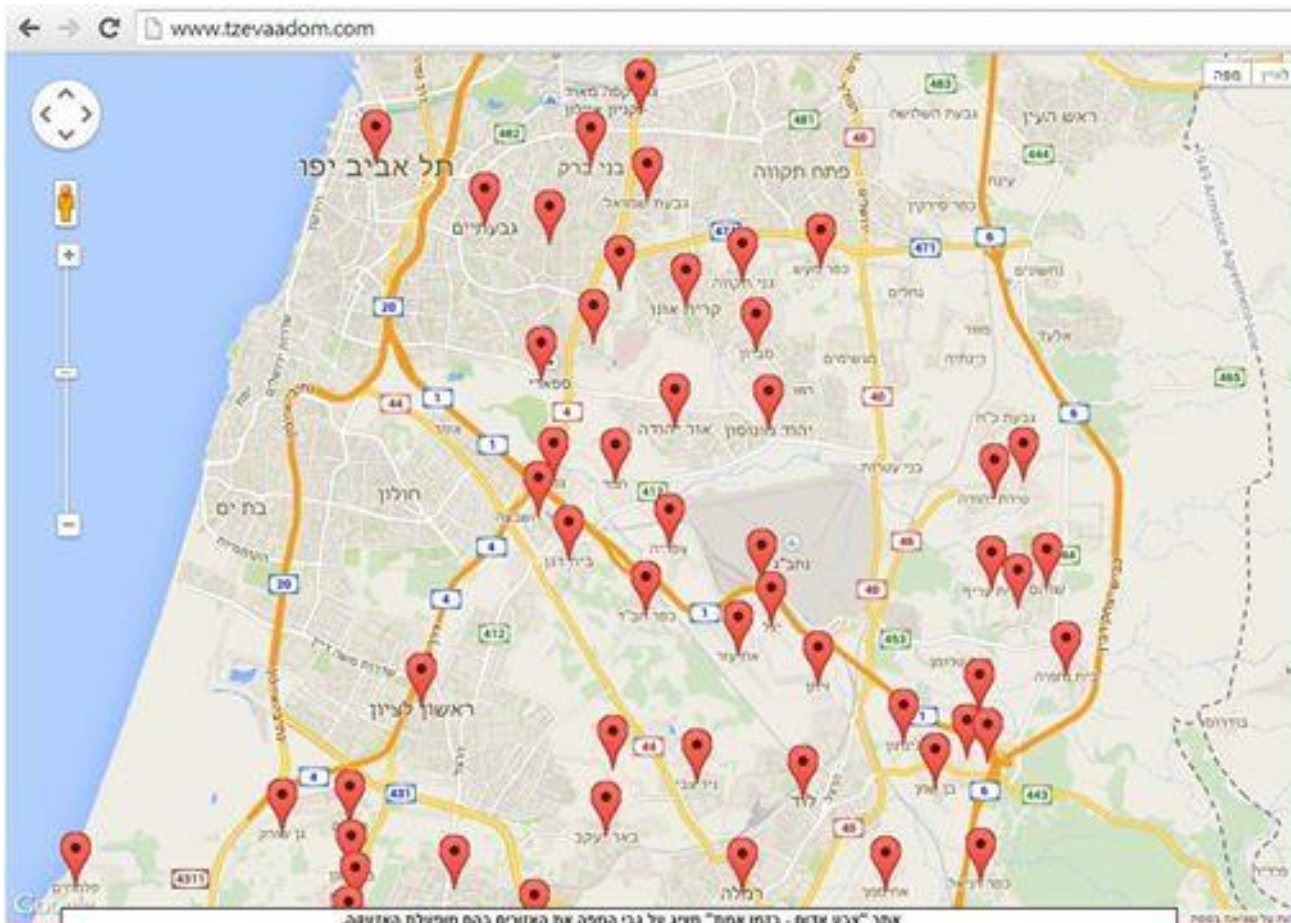
WAZE



OPENSTREETMAPS



TZEVAADOM.COM



צבע אדום

• שפלה 179

- אחיסמן
- בן שמן
- בן שמן כפר נוער
- גינתון
- כרם בן שמן
- לוד
- ניר צבי
- תעשיון צריפין

• שפלה 182

- אירוס
- גבעת ברנר
- גינתון
- גן שלמה
- כפר דניאל
- מפעלי ברנר
- נצר סרני
- רחובות

• שפלה 178

- באר יעקב
- ישרש
- מצליח
- פלמחים
- צריפין

אתר "צבע אדום - בזמן אמת" מציג על גבי המפה את האזורים בהם מפעלות האזעקה.
גרסאות beta - אתר זה אינו מהווה תחליף להוראות מאגורות פיקוד העורף

בתוך תורה ליל האזורים והתברר שישו אלינו בעיקר עזרה כלשהיית השירותים פנות , הצעת תרומות על באגים: tzevaadom@yahoo.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



OASSIS
Ontology Assisted Crowd Mining

CONTRIBUTE

STATISTICS

SEARCH

HELP

Hello SlavaNov!

1



| next star in 3 questions!

How often do you...

go to a place

2



eat at a restaurant



and also

Never

Rare

Sometimes

Often

Very often

4

Specify a place

Specify a restaurant

Submit

3

(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 fr 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!
Asking the Right Questions

AskIt! Admin View Statistics Logout

5/10


DBLP

Hosagrahar V Jagadish
University of Michigan

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



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

The screenshot shows the IBM Research Guess game interface. At the top left, it says "IBM research guess" and "1 player currently online". At the top right, it says "Welcome Ohad Greenshpan / About / Feedback / exit". The main area features a central blue circle labeled "Capital Cities" with several smaller red circles connected to it, each containing a country and its capital: "Belarus, Minsk", "Fiji, Suva", "China, Beijing", "Italy, Rome", "Japan, Tokyo", and "France, Paris". On the left side, there is a sidebar with the following information: "From Answer: +0", "Time Remaining 0:15", "Total Points 111270", and "Overall Ranking 2". Below this sidebar, it says "You are the first one to mention this name". At the bottom, there is a text input field containing "Singapore, Singapore", a green "submit" button, and a yellow "next category" button. At the very bottom, there is a "Suggestions" section with the text "Try to guess names that you think a few people guessed before to maximize your points." and a "view rules / suggest new category" link.

IBM research **guess** 1 player currently online

Welcome Ohad Greenshpan / About / Feedback / exit

From Answer: **+0**

Time Remaining **0:15**

Total Points **111270**

Overall Ranking **2**

You are the first one to mention this name

Guess as many names of **Capital Cities** as you can.

Suggestions Try to guess names that you think a few people guessed before to maximize your points. [view rules / suggest new category](#)

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

amazonmechanical turk
beta Artificial Intelligence

[Your Account](#) [HITs](#) [Qualifications](#) **110,288 HITs** available now [Sign In](#)

All HITs | [HITs Available To You](#) | [HITs Assigned To You](#)

Search for containing that pay at least \$ for which you are qualified

All HITs

1-10 of 1895 Results

Sort by: [Show all details](#) | [Hide all details](#) [1](#) [2](#) [3](#) [4](#) [5](#) > [Next](#) >> [Last](#)

| | |
|---|--|
| Image Tagging - Answer questions about ONE image. Great images! View a HIT in this group | |
| Requester: TagCow | HIT Expiration Date: Oct 24, 2010 (2 weeks 5 days) Reward: \$0.02 |
| Time Allotted: 20 minutes | HITs Available: 14019 |
| Find Restaurant Web Addresses View a HIT in this group | |
| Requester: Dolores Labs | HIT Expiration Date: Oct 12, 2010 (6 days 23 hours) Reward: \$0.07 |
| Time Allotted: 60 minutes | HITs Available: 8773 |
| Product Search Relevance View a HIT in this group | |
| Requester: Amazon Requester Inc. | HIT Expiration Date: Oct 6, 2010 (1 day 21 hours) Reward: \$0.01 |
| Time Allotted: 10 minutes | HITs Available: 7867 |
| Verify Restaurant Websites View a HIT in this group | |
| Requester: Dolores Labs | HIT Expiration Date: Oct 11, 2010 (6 days 23 hours) Reward: \$0.05 |
| Time Allotted: 60 minutes | HITs Available: 6760 |
| Find Business Web Addresses View a HIT in this group | |
| Requester: Dolores Labs | HIT Expiration Date: Oct 11, 2010 (6 days 22 hours) Reward: \$0.07 |

CROWDFLOWER

Home > Tasks > Complete Tasks My Tasks History

Tasks available

- earn
\$0.03
1599
available

[Find a Directory Listing for a Business!](#)
In this task you'll help us find a directory webpage for a variety of businesses.
- earn
\$0.03
7626
available

[Does Another Company Own This Company?](#)
Thanks very much for accepting this task! Your goal in this task is to help us determine if the business provided is owned by another company. This will be very helpful in terms of identifying relationships between different companies.
- earn
\$0.03
2748
available

[Find Descriptions of Businesses on Any Website \(Intl\)](#)
In this task we need your help finding websites that have descriptions or keywords that describe what that company does.
- earn
\$0.03
7114
available

[Categorize addresses as Residential \(Homes\) or Commercial \(Businesses\) \(US\)](#)
In this task you will look at an address and a satellite map of that address.
- earn
\$0.03
available

[Find Address Info On a Website](#)

Tasks

Earn
14
Cents
351 available

[Categorizzare servizi e prodotti offerti da un'azienda ufficiale**.](#)
In questa attività vi verrà fornito il ****nome**** di un'azienda ufficiale**.

Earn
10
Cents
1180 available

[Encontrar y Arreglar Direcciones de Negocios en Resumen](#)

Earn
1
Cents
24047 available

[Help us scrape article authors!](#)
Overview

Earn
8
Cents
825 available

[Trouver et fixer les adresses d'entreprises en France](#)
Vue d'ensemble

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

Job 575381
Finished

- 1. DESIGN JOB
 - Data
 - Build Job
 - Preview
- 2. MANAGE QUALITY
 - Test Questions
 - Contributors
 - Job Settings
- 3. GET RESULTS
 - Launch
 - Monitor
 - Results

Map Analysis: Help us Figure out the Latitude and Longitude of State and National Parks ⓘ

Seth Teicher

Dashboard Advanced Analytics

100%
Complete

\$275
Cost ⓘ

40
Active Test Questions

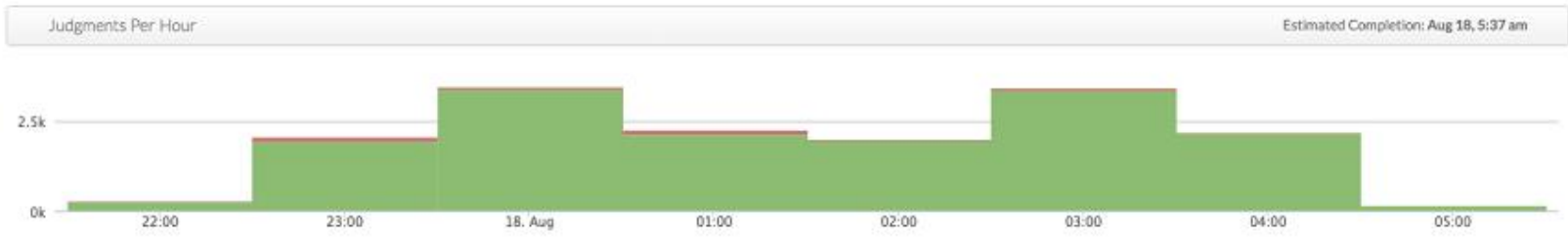
3,810
Units ⓘ

0
Judgments Per Hour ⓘ

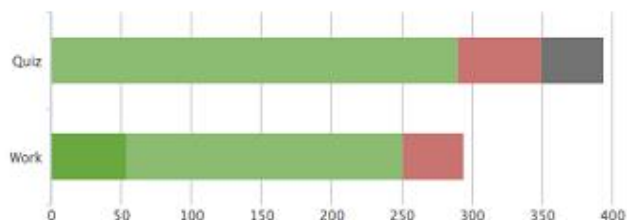
15,364
Trusted Judgments ⓘ

485
Untrusted Judgments ⓘ

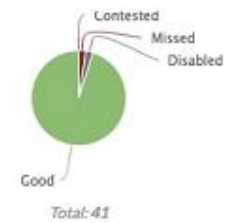
0
Pending Judgments



Contributor Funnel ⓘ



Test Questions ⓘ



Contributor Satisfaction

3.9 / 5
Overall

4.1 / 5 Instructions Clear
3.7 / 5 Ease Of Job

3.5 / 5 Test Questions Fair
3.6 / 5 Pay

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?