Active Learning and Crowd-Sourcing for Machine Translation

Vamshi Ambati, Stephan Vogel, & Jamie Carbonell

Pramod Thammaiah
Scott Brinker
November 17, 2010
CS 286r
• **Active Learning**: a few labeled instances, a large set of unlabeled instances, and a ranking of instances for an external oracle to label them

• **Active Crowd Translation**: using crowd-sourced experts and non-experts to translate sentences as the external oracle

• **Mechanical Turk**: Amazon’s crowd-sourcing platform where “requesters” post HITs (human intelligence tasks) for “turkers” to complete in exchange for micropayment rewards.
Translate Spanish to English

Instructions

- You need to be a fluent speaker of both 'Spanish' and 'English'
- For this task, do NOT use online Translation systems like Google or Babblefish.
- If you are found to do the above, your work will be REJECTED hurting your reputation index on Am
- Any words that you do not understand should be copy pasted to the translation as it is, but do not drop

Evaluation Criteria

- Your English translation should be as close as possible to the Spanish original sentence in meaning
- At least 90% of the words need to be translated for consideration of task completion. When in doubt provide untranslated

Spanish sentence:

$\{src\}$

Please provide English translation below:
• **Qualifications** for turkers can include:
  
  – Sufficient accuracy on a small test set
  
  – Minimum percentage of previously accepted submission
  
  – Geographic location (e.g., China for Chinese translation)
  
  – Option to reject unsatisfactory work

• **Pricing** for turkers varies, but generally inexpensive:
  
  – As low as as < $0.01 per translation
  
  – Empirical study in paper averaged $0.015/translation
  
  – Supply/demand factors for less common languages
Algorithm 1 Active Learning for SMT

1: Given Labeled Data Set: \( L_0 \)
2: Given Unlabeled Data Set: \( U_0 \)
3: \textbf{for} \( k = 0 \) to \( T \) \textbf{do}
4: \hspace{1em} \textbf{for} \( i = 0 \) to \( N \) \textbf{do}
5: \hspace{2em} \( s_i = \text{Query}(U_i, L_i) \)
6: \hspace{2em} Request Human Translations for \( s_i \)
7: \hspace{2em} \( S_k = S_k \cup s_i \)
8: \hspace{1em} \textbf{end for}
9: \hspace{1em} \( U_{k+1} = U_k - S_k \)
10: \hspace{1em} \( L_{k+1} = L_k \cup S_k \)
11: \hspace{1em} Re-train MT system on \( L_{k+1} \)
12: \textbf{end for}
Density Weighted Diversity Sampling (DWDS) Strategy

\[ d(S) = \frac{\sum_{x \in Phrases(S)} P(x/U) e^{-\lambda \cdot \text{count}(x/L)}}{\|Phrases(S)\|} \]  \hspace{1cm} (1)

\[ u(S) = \frac{\sum_{x \in Phrases(S)} \alpha}{\|Phrases(S)\|} \]  \hspace{1cm} (2)

\[ \alpha = \begin{cases} 1 & x \notin Phrases(L) \\ 0 & \end{cases} \]  \hspace{1cm} (3)

\[ \text{Score}(S) = \frac{(1+\beta^2)d(S)u(S)}{\beta^2 d(S) + u(S)} \]  \hspace{1cm} (4)
• **Translation Reliability Estimation** as inter-annotator agreement
  – Agreement of 3/3 translators 21.1% of the time
  – Agreement of 2/3 translators 23.8% of the time

• **Translator Reliability Estimation** as identifying reliable translators over a series of translations

\[
rel(w_k) = \sum_{t_k^i \in T_k} \sum_{t_n^j \in T} \delta(t_k^i, t_n^j) \frac{\|T_k\|}{\|T\|}
\]

\[
\delta(t_k^i, t_n^j) = \begin{cases} 
1 & t_k^i \equiv t_n^j \\
0 & \text{otherwise}
\end{cases}
\]
<table>
<thead>
<tr>
<th>System</th>
<th>Iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>crowd pick-rand</td>
<td>16.43</td>
</tr>
<tr>
<td>crowd translation-agreement</td>
<td>18.92</td>
</tr>
<tr>
<td>crowd translator-agreement</td>
<td>19.20</td>
</tr>
<tr>
<td>expert translations</td>
<td>19.21</td>
</tr>
<tr>
<td>crowd all-three</td>
<td>19.62</td>
</tr>
</tbody>
</table>
Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce.
Workers select from thousands of tasks and work whenever it’s convenient.

52,721 HITs available. View them now.

Make Money by working on HITs
HITs - Human Intelligence Tasks - are individual tasks that you work on. Find HITs now.

As a Mechanical Turk Worker you:
• Can work from home
• Choose your own work hours
• Get paid for doing good work

Find an interesting task

Work

Earn money

or learn more about being a Worker

Get Results from Mechanical Turk Workers
Ask workers to complete HITs - Human Intelligence Tasks - and get results using Mechanical Turk. Register Now

As a Mechanical Turk Requester you:
• Have access to a global, on-demand, 24 x 7 workforce
• Get thousands of HITs completed in minutes
• Pay only when you’re satisfied with the results

Funding your account

Load your tasks

Get results

Get Started
<table>
<thead>
<tr>
<th>HIT Title</th>
<th>Requester</th>
<th>HIT Expiration Date</th>
<th>Reward</th>
<th>HITs Available</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tag One Photo</td>
<td>Roger Guzman</td>
<td>Nov 18, 2010 (3 days 3 hours)</td>
<td>$0.01</td>
<td>4332</td>
</tr>
<tr>
<td>Translate Malayalam words into English</td>
<td>Chris Caliion-Burch</td>
<td>Nov 22, 2010 (6 days 23 hours)</td>
<td>$0.05</td>
<td>4092</td>
</tr>
<tr>
<td>Find the email address for the company and website</td>
<td>Sam GONZALES</td>
<td>Nov 24, 2010 (1 week 1 day)</td>
<td>$0.01</td>
<td>3719</td>
</tr>
<tr>
<td>上駄金案の提出書類の確認</td>
<td>Goodjob</td>
<td>Nov 18, 2010 (3 days)</td>
<td>$0.05</td>
<td>3650</td>
</tr>
<tr>
<td>Find Restaurant Web Addresses I</td>
<td>Dolores Labs</td>
<td>Nov 22, 2010 (6 days 23 hours)</td>
<td>$0.05</td>
<td>3615</td>
</tr>
<tr>
<td>Find Product Information for Sporting Apparel</td>
<td>Dolores Labs</td>
<td>Nov 19, 2010 (4 days 12 hours)</td>
<td>$0.12</td>
<td>2897</td>
</tr>
<tr>
<td>Rewrite, end answer a cooking related question</td>
<td>Mork Foodista</td>
<td>Nov 24, 2010 (1 week 2 days)</td>
<td>$0.05</td>
<td>1887</td>
</tr>
<tr>
<td>Rate Images for Quality and Relevance</td>
<td>Dolores Labs</td>
<td>Nov 22, 2010 (6 days 23 hours)</td>
<td>$0.15</td>
<td>1823</td>
</tr>
<tr>
<td>CopyEditing and Logically Filling up of Blanks for Recipe Database</td>
<td>Arthur</td>
<td>Feb 1, 2011 (11 weeks 1 day)</td>
<td>$0.40</td>
<td>1400</td>
</tr>
</tbody>
</table>
Mechanical Turk in Commercial Applications

- Catalog and data management
  - Confirm accuracy of catalog data, identify duplicates
  - Select best images to showcase a catalog item

- Database creation
  - Content harvesting

- Search optimization & content management
  - Tag content with keywords to improve searchability
  - Ensure content adheres to certain guidelines
Questions About Crowdsourcing with Mechanical Turk

• Other applications?

• Ethical concerns?
Crowdsourcing and All-Pay Auctions

Pramod Thammaiah
Scott Brinker
November 17, 2010
CS 286r
Introduction

• Examines crowdsourcing in things like Tasken, Yahoo! Answers, etc.
• Want to understand the relationship between rewards and participation rates
• Presents mathematical model and empirical analysis based off of All-pay auctions
## Logo Design Contests

Our team of logo design professionals compete to produce a unique logo design for you. Post a contest, review the designs and pick a winner, it's that easy.

### Contest Title
- **Dog Rescue Trying to Play With The Big Dogs**
  - We need to re-brand ourselves starting with a sophisticated logo...
  - **Maribel1**
  - Ends: 2 days, 4 hrs
  - Entries: 581
  - Prize: $1,500

- **MEDITERRANEAN RESTAURANT LOGO RE-DESIGN**
  - We are a local 3 location Mediterranean restaurant with a passion for good food...
  - **Ugan**
  - Ends: 5 days, 21 hrs
  - Entries: 77
  - Prize: $1,000

- **Private Contest**
  - Login to view...
  - **Elite Unit**
  - Ends: 3 days, 20 hrs
  - Entries: 20
  - Prize: $1,000

- **Looking for an elegant, crisp, cool logo**
  - Conway Wealth Group is a service team at Summit Financial...
  - **CWG**
  - Ends: 5 days, 3 hrs
  - Entries: 768
  - Prize: $1,000

- **Social Stork Facebook App Logo**
  - Social Stork is a new social moms Facebook application al...
  - **Keith Seim**
  - Ends: 2 days, 9 hrs
  - Entries: 160
  - Prize: $550

- **NEW!! Logo Design for Australian Industrial Equipment Supplier**
  - Pacific Automation is a distributor of industrial product...
  - **Sbryant1**
  - Ends: 2 days, 14 hrs
  - Entries: 119
  - Prize: $550

- **Logo Design for Fantasy Sports Company**
  - Playerline is a fantasy sports company that provides the ...  
  - **Playerline**
  - Ends: 8 hrs, 22 mins
  - Entries: 1101
  - Prize: $500

- **IAccel Productivity Suite - needs a great logo and identity**
  - IAccel is a suite of Productivity Tools for Professional ...  
  - **Rubicon Software**
  - Ends: 2 days, 4 hrs
  - Entries: 57
  - Prize: $500

- **blind**
  - We need a logo for a consumer electronics brand...
  - **smartparts**
  - Ends: 2 days, 7 hrs
  - Entries: 303
  - Prize: $591

---

**Categories**
- **all categories** (583)
- web page design (85)
- stationery design (11)
- logo design (580)
- button & icon design (10)
- other graphic design (29)
- t-shirt design (11)
- banner ad design (4)
- print design (27)
- twitter background design (1)
- quickbooks form design (0)
- custom wordpress themes (14)

**Search Filters**
- Keywords
- Contest ends
- Min prize
- Guaranteed
- Only blind

**Recent Winning Designers**
Dog Rescue Trying to Play With The Big Dogs

Entries highest rated first

#483 By igno ★★★★★
#472 By Maestral ★★★★★
#471 By secondlui ★★★★★
#470 By Maestral ★★★★★
#469 By secondlui ★★★★★

#465 By Maestral ★★★★★
#464 By Maestral ★★★★★
#463 By Maestral ★★★★★
#462 By Maestral ★★★★★
#461 By Maestral ★★★★★

#460 By Maestral ★★★★★
#378 By Dalibor. ★★★★★
#351 By Dalibor. ★★★★★
#240 By umbre ★★★★★
#133 By Andrew ★★★★★
All Pay Auctions

- All-pay auctions are those where each agent pays their bid before allocation of the good.

- The highest bidder wins the good.

- Examples: political elections, Swoopo, lobbying, bidding on the value of a wallet, etc.
Basic Model

• Consider a 2-stage all-pay auction
  – Each player selects a contest and makes a bid (think of the bid as effort)
  – For each contest, the player with the highest bid wins

• Each player has a private skill that is known only to them

• The reward of each contest and the distribution of skills is known to all players
Mock contest

• Pick one of the 2 contests:
  – Write the best joke
  – Write the best riddle

• The winner will be selected by Prof. Chen
  – So make them short....

• The best joke will get $1 and the best riddle will get $2
Mathematical Formulation

• There are $N$ players that choose among $J$ contests
• Let $R_j$ denote the reward offered in contest $j \in \{1, \cdots, J\}$
• Each player $i$ has contest specific skills $\bar{v}_i = (v_{i1}, \cdots, v_{iJ})$ drawn independently from continuous joint probability distribution over $[0, m]^J$
• The skills of each player are assumed to be independent. Players do not know other players skills, but do know the distribution
• The game has 2 stages:
  – In stage 1, player $i$ selects a contest $j$ and a bid $b_{ij}$
  – In stage 2, the player with the highest bid is selected for each contest. A winning player has payoff $v_{ij} R_j - b_{ij}$ and a losing player has payoff $-b_{ij}$
• A mixed strategy consists of a probability distribution $\bar{\pi}_i = (\pi_{i1}, \cdots, \pi_{iJ})$ a bid for each contest
• Equilibrium concept is Bayes-Nash equilibrium
• Assume that bids are strictly monotonically increasing with skill levels so that the winner is the player with the highest skill level
Proposition 3.1: There exists a symmetric equilibrium

- Let $p_j$ be the probability that a player selects contest $j$, $\hat{F}_j(v)$ be the cumulative distribution over skill for a player given that he selects $j$, and $\hat{F}^c_j(v) = 1 - \hat{F}_j(v)$

- Let $g_j(v)$ denote the expected profit of a player with skill $v$ for contest $j$ then using Revenue Equivalence we have:

$$g_j(v) = R_j \int_0^v (1 - p_j \hat{F}^c_j(x))^{N-1} dx$$

- Note that this is not an unique equilibrium
  - Consider 2 players and 2 contests. Then player 1 always picking contest 1 and player 2 always picking contest 2 regardless of skill is an asymmetric equilibrium

- Corollary: Given a set of contests with the same reward. A player will only choose a contest in that he has the maximum skill.

- Assume that all players only choose symmetric strategies
Large-System Limit

- The number of contests needs to stay proportional to the number of agents in the limit
- Assume that there are only, $K$, finitely many classes of rewards
- The number of participants in each contest is a Poisson random variable, whose mean is logarithmic in the size of the reward
Player-Specific skills

• Assume that every contest requires the same skill, or formally, $\vec{v}_i$ is equal to $(v, v, \ldots, v)$

• In this case the symmetric equilibrium is unique.

• Let contests be grouped into $K$ classes, based on having the same reward
  – Where $R_1 > R_2 > \ldots > R_K$ and for notational simplicity: $\vec{R} = (R_1, \ldots, R_K)$
  – For any subset $A \subseteq \{1, \ldots, K\}$ we define:

$$H_A(\vec{R}) = \left( \sum_{k \in A} \frac{J_k}{J_A} R_k^{-\frac{1}{N-1}} \right)^{-1}$$

$$J_A = \sum_{k \in A} J_k.$$
Theorem 4.1

• Under the player-specific skills, the symmetric equilibrium satisfies the following two properties:

1. **Threshold Reward:** A contest is selected by a player with strictly positive probability only if the reward offered by this contest is one of the $\tilde{K}$ highest rewards, where

$$\tilde{K} = \max \left\{ i : R_i^{N-1} > \left( 1 - \frac{1}{j_{[1,i]}} \right) H_{[1,i]}(\bar{R}) \right\}$$

   — Intuition: At a certain point, contests with low rewards will get no participants.

2. **Participation rates:** A player selects a particular contest of class $j$ with probability given by $p_j$

$$p_j = \begin{cases} 
1 - \left( 1 - \frac{1}{j_{[1,\tilde{K}]}} \right) \frac{H_{[1,\tilde{K}]}(\bar{R})}{R_j^{N-1}}, & \text{if } j \leq \tilde{K} \\
0, & \text{otherwise}
\end{cases}$$
Players have a minimum reward level and compete in contests at or above this level with decreasing probability. The minimum reward level increases with skill level. Overall, contests with higher rewards get more players.
Contest-Specific Skills

- A player’s skills for each contest are drawn independently.
- A player will only need to pay attention to his highest skill in each reward class.
Extensions

• Asymmetric Skills
  – Suppose the maximum skill $m$ is specific to each contest class
  – Can we infer these maximal skill levels by examining participation levels? (ie. Make it endogenous)

• Minimum Effort
  – Corresponds to having a minimum bid, or entry fee
System Design

• Maximize overall social welfare
• There is utility gained with a greater number of participants in a contest.
• There is a cost associated with the rewards that are paid
• In a zero-cost setting, optimal rewards are unique up to a multiplicative constant
All-Pay Auctions Empirical Results (taskcn.com)
Comparing Empirical Analysis to Analytical Predictions

• Tasks on the Chinese web site Taskcn
• Tasks with a single winner in the year 2008
• Graphics, Characters, and Miscellaneous categories
• Particular focus on Graphics category including:
  – Logos subcategory
  – 2-D design subcategory
  – Conjecture: these subcategories use homogeneous skills
<table>
<thead>
<tr>
<th>Category</th>
<th>Number of tasks</th>
<th>Submissions per task</th>
<th>Median submission period (days)</th>
<th>Median reward (RMB)</th>
<th>Median total submissions per user (observed on Feb ’09)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graphics</td>
<td>2,392</td>
<td>44.66</td>
<td>16.97</td>
<td>200</td>
<td>n/a</td>
</tr>
<tr>
<td>Logos</td>
<td>571</td>
<td>50.36</td>
<td>16.96</td>
<td>240</td>
<td>24</td>
</tr>
<tr>
<td>2-D</td>
<td>1,431</td>
<td>49.14</td>
<td>17.68</td>
<td>200</td>
<td>23</td>
</tr>
<tr>
<td>Characters</td>
<td>613</td>
<td>323.75</td>
<td>17.15</td>
<td>80</td>
<td>n/a</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td>565</td>
<td>38.47</td>
<td>12.29</td>
<td>79.85</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Table 1: Summary of dataset basic properties.

(i) Graphics  
(ii) Characters  
(iii) Miscellaneous

Sort of logarithmic  
Heterogeneity of tasks?
(i) Graphics

(ii) Characters
1,431 tasks had 774 distinct winners—however, a group of 122 users accounted for 50% of all winning submissions
Summary

• Players randomize dependent on the skill level they get
• Contests with rewards below a certain threshold get no participants
• Players in certain intervals of skills have a certain interval of contests they participate in
  – As skills increase, they only participate in the highest paying contest
• Logarithmic relationship between rewards and participation (in large systems)
Further Questions

• Why can players only pick one contest?
  – Not the case since 122 players won over 50% of 1431 contests
• What if we had multiple winners?
• How can we capture the “time-element” of these situations?
• What if the market becomes so big that it becomes difficult to search for tasks?
• Why is there a drop-off in participation for contests with high rewards?
• What may be other highly influential externalities?
• Differences between money and reputation?
<table>
<thead>
<tr>
<th>Vote</th>
<th>Answer</th>
<th>Views</th>
<th>Title</th>
<th>Score</th>
<th>User</th>
<th>Date</th>
<th>Score</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>3</td>
<td>93</td>
<td>iphone 4 - Application doesn't show image?</td>
<td>+50</td>
<td>levis501</td>
<td>1d ago</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>170</td>
<td>Debugging SimpleCursorAdapter</td>
<td>+200</td>
<td>Emile</td>
<td>16h ago</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>56</td>
<td>15</td>
<td>3</td>
<td>The Definitive C Book Guide and List</td>
<td>+200</td>
<td>Ben Jackson</td>
<td>2d ago</td>
<td>2,713</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>120</td>
<td>Understanding Fusion Trees?</td>
<td>+100</td>
<td>belisarius</td>
<td>nov 4</td>
<td>5,372</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>78</td>
<td>2d Light effect with SDL</td>
<td>+50</td>
<td>jv42</td>
<td>6h ago</td>
<td>1,534</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>36</td>
<td>Debug: Need StackTrace on mono via mod_mono+apache2</td>
<td>+50</td>
<td>acidzombie24</td>
<td>nov 2</td>
<td>6,625</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>375</td>
<td>Reusing validation attributes in custom ViewModel</td>
<td>+100</td>
<td>jirar</td>
<td>nov 1</td>
<td>7,668</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>67</td>
<td>App_offline.htm created/deleted whenever I check out a file</td>
<td>+50</td>
<td>XwineoutX</td>
<td>oct 31</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>134</td>
<td>How does one define PATH_MAX for a given filesystem?</td>
<td>+100</td>
<td>Matt Joiner</td>
<td>nov 1</td>
<td>7,726</td>
<td></td>
</tr>
</tbody>
</table>
I've been wondering whether there is a good "git export" solution that creates a copy of a tree without the .git repository directory. There are at least three methods I know of:

1. `git clone` followed by removing the .git repository directory.
2. `git checkout-index` alludes to this functionality but starts with "Just read the desired tree into the index..." which I'm not entirely sure how to do.
3. `git-export` is a third party script that essentially does a git clone into a temporary location followed by `rsync --exclude='.git'` into the final destination.

None of these solutions really strike me as being satisfactory. The closest one to `svn export` might be option 1, because both those require the target directory to be empty first. But option 2 seems even better, assuming I can figure out what it means to read a tree into the index.

2. I think you should consider accepting the `git-archivar-answer`. -- Josef Dec 1 '09 at 22:58

---

14 Answers

Probably the simplest way to achieve this is with `git archive`. If you really need just the expanded tree you can do something like this.

```
# tar will pack to a file

# tar will pack to a directory
```

Most of the time that I need to 'export' something from git, I want a compressed archive in any case so I do something like this.

```
# bzip2 compresses

# gzip compresses
```

`git help archive` for more details, it's quite flexible.
I found out what option 2 means. From a repository, you can do:

```
git checkout-index -a -f --prefix=/destination/path/
```

Since in a normal situation the index contains the contents of the repository, there is nothing special to do to "read the desired tree into the index". It's already there.

The `-a` flag is required to check out all files in the index (I'm not sure what it means to omit this flag in this situation, since it doesn't do what I want). The `-f` flag forces overwriting any existing files in the output, which this command doesn't normally do.

This appears to be the sort of "git export" I was looking for.

```
...and DON'T FORGET THE SLASH AT THE END, or you won't have the desired effect :) – conny Apr 8 '09 at 20:48
```

Isn't the index just a name for the "staging" area? How's it that the tree is already there? I thought you can only add stuff manually to it with `git add` – hasen Aug 31 '09 at 6:12

The `git add` command changes content in the index, so whatever `git status` shows as "to be committed" is the differences between HEAD and the contents of the index. – Greg Hewgill Aug 31 '09 at 6:30

seems that this is equivalent to `git archive HEAD` – Takeshin May 5 at 20:33

@conny: read your comment, forgot about it and ran the command without a trailing slash. tip: follow conny's advice. -- Znarkus Jun 24 at 16:55

`git archive` also works with remote repository.

```
git archive --format=tar --remote=ssh://remote_server/remote_repository master
```

Disabling or overriding the excel and pdf export function in Reporting
<table>
<thead>
<tr>
<th>Rank</th>
<th>Username</th>
<th>#</th>
<th>Change</th>
<th>Total Reputation</th>
<th>Monthly Reputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>Nick Craver</td>
<td>0</td>
<td>104,771</td>
<td>2,625</td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td>Hans Passant</td>
<td>+1</td>
<td>109,727</td>
<td>1,965</td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td>Pascal Thivent</td>
<td>+2</td>
<td>109,176</td>
<td>1,815</td>
<td></td>
</tr>
<tr>
<td>#4</td>
<td>Jon Skeet</td>
<td>-2</td>
<td>235,008</td>
<td>1,786</td>
<td></td>
</tr>
<tr>
<td>#5</td>
<td>Darin Dimitrov</td>
<td>-1</td>
<td>113,508</td>
<td>1,725</td>
<td></td>
</tr>
<tr>
<td>#6</td>
<td>Marc Gravell</td>
<td>+11</td>
<td>175,255</td>
<td>1,663</td>
<td></td>
</tr>
<tr>
<td>#7</td>
<td>thejh</td>
<td>+1445</td>
<td>2,046</td>
<td>1,642</td>
<td></td>
</tr>
</tbody>
</table>

When your fellow users vote up your questions and answers on a Stack Exchange site, you generate reputation. Reputation is a rough measure of:
- how much the community trusts you
- your communication skills
- the quality and relevancy of your questions and answers

These friendly **reputation leagues** are an informal way of tracking your reputation within the community on a particular Stack Exchange.

Reputation is capped at **200** per day, but remember that bounty awards and accepted answers are immune to this daily reputation cap.

### Top New Users

- **Hansmukh**
  - 402 reputation this month
  - Joined: Nov 2, 2010
- **Surreal Dreams**
  - 525 reputation this month
  - Joined: Nov 3, 2010
- **zzzzzBov**
  - 392 reputation this month
  - Joined: Nov 4, 2010
- **Altruist** × 211: First bounty you manually awarded on another person's question.
- **Announcer** × 61: Shared a link to a question that was visited by 25 unique IP addresses in 3 days.
- **Autobiographer** × 35439: Completed all user profile fields.
- **Benefactor** × 4005: First bounty you manually awarded on your own question.
- **Beta** × 2586: Actively participated in the Stack Overflow private beta.
- **Booster** × 8: Shared a link to a question that was visited by 300 unique IP addresses in 4 days.
- **Citizen Patrol** × 7438: First flagged post.
- **Civic Duty** × 6043: Voted 300 or more times.
- **Cleanup** × 4841: First rollback.
- **Commentator** × 44678: Left 10 comments.
- **Copy Editor** × 89: Edited 600 entries.
- **Critic** × 31920: First down vote.
- **Disciplined** × 2030: Deleted own post with score of 3 or higher.
- **Editor** × 92035: First edit.
- **Electorate** × 522: Voted on 600 questions and 25% or more of total votes are on questions.
- **Enlightened** × 15571: First answer was accepted with score of 10 or more.
- **Enthusiast** × 9558: Visited the site each day for 30 consecutive days.
- **Epic** × 113: Hit the daily reputation cap on 50 days.
- **Famous Question** × 2776: Asked a question with 10,000 views.
- **Fanatic** × 1484: Visited the site each day for 100 consecutive days.
- **Favorite Question** × 1574: Question favorited by 25 users.
- **Generalist** × 164: Provided non-wiki answers of 15 total score in 20 of top 40 tags.
- **Good Answer** × 11949: Answer score of 25 or more.
- **Good Question** × 3446: Question score of 25 or more.