QuestionMarket

A marketplace mechanism for tapping into the value of human cognition

Benjamin Hill
UC Berkeley / SIMS

Abstract

A *QuestionMarket* enables the exchange of units of human-generated information. While many other distributed 'grid' systems create a marketplace for trading computational resources like processing, bandwidth, and storage, QuestionMarket deals in the acquisition and exchange of small units of human cognition. By focusing on puzzles that are difficult to solve with a computer but relatively easy for humans, so called AI-complete problems are resistant to Moore's Law while encompassing a wide range of valuable human-only work, such as translation, tagging, and filtering. Building on the existing multi-billion-dollar market in trading very fine-grained units of human attention (advertising), QuestionMarket lies at the intersection of web-based advertising, micro-outsourcing of human work, multimedia metadata annotation, and focus group testing. Most intriguingly, market-based reward systems -- prices -- should serve to direct the right puzzles to the right solvers, whether the intelligence required is human or artificial. By putting a price on answers, QuestionMarket posits a fair playing field for benchmarking future improvements in artificial intelligence against typical human performance.

ACM Keywords

C.2.4 Distributed Systems - Distributed applications

H.5.2 User Interfaces

H.5.3 Group and Organization Interface - Computer-supported cooperative work

I.2.11 Distributed Artificial Intelligence

I.4.8 Scene Analysis

K.8 PERSONAL COMPUTING

General Keywords

Distributed Human Computation

Web Advertising

Multimedia annotation

Economics

Assisted AI

1. The Challenge

Today, humans still outperform even the best Artificial Intelligence ("AI") systems for a wide variety of perceptual tasks. While AI continues to make incremental improvements, the term AI-Complete (*http://en.wikipedia.org/wiki/AI-complete*) label for the sorts of problems that are difficult to program solutions to, yet relatively easy for people to answer (or at least offer an opinion on). For some of these AI-Complete tasks, QuestionMarket proposes a standard format for presenting multimedia puzzles to users, a market mechanism for incentivizing users, means of matching problem to solver to optimize efficiency, and correlating/verifying the user generated answers to increase data quality.

Table 1: *Examples of familiar AI-Complete challenges that QuestionMarket can address:*

Is this image inappropriate for children under 17?
Annotate this image with descriptive tags
Which of these 5 search result URLs are spam or irrelevant?
Translate this text into an uncommon language.
Rate this face from 1 to 10
What search terms would you use to find/describe this blog?
What is the text in this captcha image?
How much do you like this clip from a new pop song?
Which of these two logos are better for this product?
Which hairstyle makes you trust this politician more?
Do you think this commercial would make increase sales?
What age range do you think would like this song the most?
Do you recognize any celebrities in these pictures?

An efficient QuestionMarket solution would require the following features:

- A question-answer data model that is flexible enough to meet a wide range of questions, while keeping the answer format simple enough to validate and consolidate.

- A display model that is able to leverage existing Web advertising as a "distribution channel."

- An interaction model that matches expected and familiar web interaction patterns

- A profiling method that builds on common Web surfers' skills, while matching specialized skill sets with high-value questions

- For launch, an existing community of users that combines a need, a community of users, and a low transaction-cost incentive mechanism

2. Background and Related Work

QuestionMarket grew out of an investigation into market-style control of peer-to-peer systems, as initially discussed in the [PeerMarket working paper](). It drew upon related work in grid computing and framed the challenge in terms of finding a common model that would enable trading traditional units of computational power. Later, as we adapted the work in light of studying Web advertising, AI, and, in particular, the relatively recent concept of *distributed human computation,* the problem statement shifted towards trading units of human attention.

2.1 Grid Computing

It has become a cliché to observe that the resources available at the "edge of the internet" have massive computational, network, and storage resources (recently combined into a novel measure of [computons]()). This collective power can tackle computer-solvable problems such as distributed backup ([OceanStore]()), download ([BitTorrent]()), protein folding ([folding@home]()), signal detection, and others. As the race to capture excess computons intensifies, these computing needs could be met by an open marketplace that better incentivizes owners to trade computing resources, while minimizing the overhead and transaction costs of assembling a coalition of willing computers.

However, a computon based problem solving marketplace may not be the area of highest potential value. All of the elements of a computon are subject to [Moore's Law](), which continually erodes the equilibrium point between the increased transaction costs of a decentralized solution and the plummeting real costs of [purchasing such resources at a central location](), with unexpected limits such as [electricity]() limiting the efficiency of very-large-scale computing. Unless the social aspects of the problem require a politically decentralized or essentially free so-

lution (as with filesharing), Moore's law may give the advantage to centralized "utility computing" when there is a single organization that needs an answer to a large problem. The uneven resources, quality, and availability that a distributed marketplace provides becomes a liability, potentially restricting a computon market's viability.

2.2   Distributed Human Computation

Distributed human computation (DHC) is currently enjoying substantial success within Web communities with shared goals. Tagging communities such as delicious and flickr take advantage of a community-wide process of resource organization by making visible the network effect generated by consistent multimedia annotation tags — the greater the consensus on the meaning of tags, the easier and more rewarding it is to share collections. Using similar network-effect motivated participants, ranking sites such as HotOrNot have a "fun" factor motivating the ranking of user images.

2.3   Artificial Intelligence

A common theme of the potential QuestionMarket table presented earlier is the class of "multimedia metadata markup" tasks, a theme that intersects with many current AI challenges. For example, even the best AI's image recognition ability rapidly decreases as it attempts to identify barcodes, printed text, manuscripts, and handwriting, while a human's ability follows a much different ability curve. The less certain an automated result is, the greater the incentive to cross-check the work by a human review panel. If a marketplace can reduce the transaction costs of soliciting, ranking, and rewarding workers to perform occasional work, an Internet-scale DHC service could be a boon to all sorts of AI research, providing a real-time incentive based ranking system in which new AI techniques could continually compete against the time-value of human effort.

2.4   Advertising

The ubiquity of Web advertising, specifically keyword auctions , reflects the viability of a marketplace for 'micropayments' for human attention. Most ads today only demand screen space and network bandwidth, placing little to no cognitive demand on the user. Advertisers are locked in a constant arms race, attempting to bring the advertising to the user's attention without overly annoying the user. At a minimum, text or banner ads fail to solicit any more mental effort than (potentially) remembering a product or deciding to click on a link. Increasingly aggressive ad formats block access to content through an enforced delay or by obscur-

ing the content until dismissed. Fully interactive models of advertising such as 'advergames' or 'surveys' appear to depend on user feedback, while in reality behave like simple click-through banners. Only a few cases exist of directly tying ad viewership to value transfers, most famously in Salon's sponsored content access ad-with-quiz solution.

3. The QuestionMarket Proposal

A QuestionMarket requires a simple, yet flexible model of the work expected of a human or machine solver. Furthermore, such an abstract model must admit comparison and summarization of "solutions," so that an automated market mechanism can expect to set price signals that attract the 'right' solvers to the 'right' problems. Our proposal is to encapsulate units of multimedia-to-metadata processing using a prompt (an URL-addressable resource presented within an HTML frameset) and an N-dimensional space where solutions are vectors of binary- or integer-magnitudes.

Our hypothesis is that such an abstraction of a DHC unit of work is 1) useful; 2) interchangeable enough to become a medium of exchange; and 3) can be delivered to end-users with transaction costs similar to that of Web advertising. Such a result would advance our overall goal of generating higher value than passive advertising — a micropayments system, if you will.

In this section, we will discuss our model of questions; and how a market in trading answers for such questions might operate.

3.1 Designing Questions

Document Engineering offers an approach to describing how organizations interact in terms of the document artifacts they exchange, rather than the processes that govern particular workflows in effect at one point in time. In this context, Web advertising today can be described as a contract offering payment to an intermediary for presenting an encapsulated multimedia unit that (hopefully) the user will pay attention to and act upon. Whether that advertisement is accepted on the basis of a second-price auction, cost-per-action, or per-view basis is a critical dimension of process, of course, but the idea that any offer must encapsulate the identity of the advertiser, the period of time the offer is valid, the content to be presented, and so on, is the core of a Document Engineering analysis of application integration.

Based on a Document Engineering analysis of web advertising, DHC, and [multimedia annotation](#) , we propose the following document model for the participants in a QuestionMarket:

Question

1. Type of problem (tag/fixed tag/rate/rank/filter/free-text)

2. Instructions (a human-readable explanation)

3. Payment terms (price and quantity parameters)

4. Ordered list of multimedia elements (the battery of text/html/image/video/audio/url prompts)

5. Optional fixed vocabulary list (the basis vectors of the desired answer format)

6. Hints about the skills required of a solver (e.g. visual/musical/social/...)

Answer

1. A list of integer or Boolean magnitudes representing an N-dimensional vector (where N is size of the tag vocabulary)

We will have to experimentally evaluate the appropriateness of this proposed model — should prices be fixed, or vary over time? It is also intriguing to consider that this model can also support "multiple-choice" questions, where a small set of preassigned vectors are offered to the user as choices. That restriction of the question's range could further be composed with a standard Boolean validation challenge: "Is answer A correct for question Q?" — that is, one could use a QuestionMarket for peer review of a sort, enabling decentralized governance.

3.2   Making a Market in Answers

Asking people to spend time and effort to actively answer questions calls for some form of compensation. With passive advertising, the deal is an implicit one: pay attention to some content you might not want in order to get access to content you do want. Making the bargain explicit is only part of the solution — there certainly are ways to reward people directly for watching advertisements. Making the reward contingent on performance is what makes this a dynamic market amongst competitive traders.

We presume that an individual solver's compensation should be based on (at least): the difficulty of the question being asked; the scarcity of the skills needed to answer the question; peer review of the correctness of the answers; and the past quality of the given user's answers (reputation). Can price signals capture all of these aspects?

An initial iteration of QuestionMarket would function like an early model for Web advertising: distributing questions to users randomly, with a fixed payoff. Less randomly, users might select the questions they want to pursue, but this has only been demonstrated with "macro-payments" (Google Answers , Experts Exchange , etc) or for free (Usenet, Wondir).

If successful, the QuestionMarket pricing structure would automatically evolve the control structures necessary to match skilled users with valuable questions, turning the pricing problem of QuestionMarket questions into a market optimization problem. These market control structures seem more applicable for QuestionMarket than for computons, a commodity where one computon/reliability combination is interchangeable with another and prices should converge.

A potential question-to-answer matching model is Jeopardy!-style questions, where a user can answer a question or choose to switch topics. A "correct" answer would increase the user's rating in that category of question, offering them higher-value questions in the future. In addition, answering questions correctly may change the compensation that an individual receives for units of their work/time.

We can already identify several challenges to the market driven model that will require additional study:

- It may be difficult to differentiate between random answers, and "local maximum" sets of answers that express the opinion of a minority.

- If a user is truly unable to answer a question, should there be a way to opt-out to avoid random guesses?

- Will this mechanism work for narrower audience with more rarefied skills? As the opportunity costs for the solver go up, would such a mechanism attract a new user population with lower costs in the developing world? Or can pattern recognition techniques tap into the wisdom of crowds to answer questions based on several less-accurate solvers' opinions?

- How significant is the distinction between providing compensation in the form of electronic tokens or cash instead of electronic goods, information services, or other "tokens" (e.g. Sennari)

- How is a "correct" answer determined?

There are many methods of forming a dynamic market environment, some of which are extremely applicable to the QuestionMarket space. Some methods require individual user profiling over time, while others would be applicable to both tracked and anonymous users. QuestionMarket plans to offer some of the following techniques:

- Incremental jumps in difficulty and compensation for "correct" answers (answers that correspond to other answers) would automatically self-calibrate users.

- Lotteries for payment, where random questions would have a higher compensation for a correct answer, could increase users' willingness to answer series of questions.

- Traditional no-cost incentive mechanisms become even more applicable: For example, users may be rewarded with seeing current results for opinion questions, but only after answering the poll themselves.

- Reputation systems (including a user ranking and meta-moderation) encourages user effort and non-random answering, allowing an otherwise anonymous 2-sided market to react to the quality of users' answers.

4.   Implementation

The current version of the QuestionMarket prototype resembles existing advertising services that serve up banner and text advertisements on participating sites. User accounts and credit balances are stored in a centralized web-based server, and users must log in to create new jobs, view job status, or answer other users' jobs. This can be easily extended to allow anonymous users to answer jobs for a specific user without logging in. This mode allows any user of the system to farm out jobs to the internet community in a method similar to banner advertising.

The prototype is implemented in PHP using PEAR DB and QuickForm. It includes user accounts, account balance, purchasing credits, and setting pricing for units of DHC. The DHC Unit creation form is nearly complete, and successfully stores a variety of multimedia element formats. Remaining tasks include building out a puzzle solving interface that is flexible

enough to handle all types of media in and vector out, while being consistent enough to end users as to look like a normal easy-to-understand HTML form.

The following examples were chosen to specifically contrast with the existing web survey methodology literature .
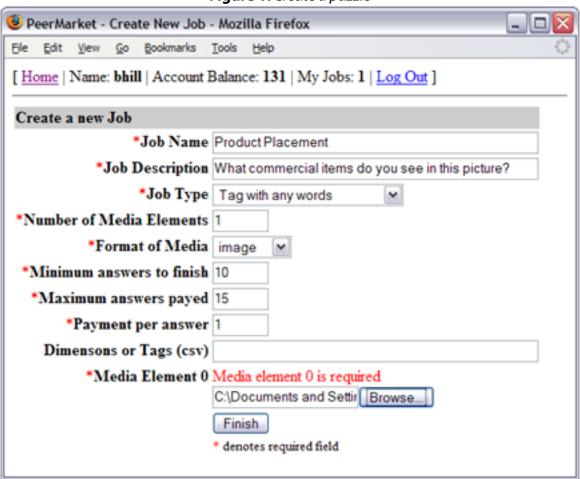
Figure 1: *Create a puzzle*

Figure 2: *Solve a puzzle (example 1)*

Figure 3: *Solve a puzzle (example 2)*

Figure 4: *Puzzle statistics and results*

4.1 Encoding Examples

The same set of potentially valuable questions is reformatted to fit within the proposed QuestionMarket model.

Table 2: *Examples of puzzle encodings*

Example Questions	Input Format	Vector-Based Answers	Skills Required
Which of these 5 URLs are spam results or irrelevant for this search term?	URL list 1..n	spam_n(0-1), relevant_n(0-1)	knowledge of spam, visual
Translate this text into an uncommon language.	Text or Image	Text	knowledge of both languages
Which of these images are not appropriate for children under 17?	Image list 1..n	under17_n(0-1)	acceptability judgement, visual

Example Questions	Input Format	Vector-Based Answers	Skills Required
Tag this image with descriptive tags	Image	unlimited unrestricted dimension tags, no scalar	visual, language specific, creativity
Rate this face from 1 to 10.	Image	hot(1-10)	visual, social
What search terms would you use to find] this website?	URL	unrestricted dimension tags, no scalar	english, prediction
What does this captcha read ?	Image	single tag, no scalar	visual, language specific
How much do you like this clip from a new pop song?	audio	like(0-10)	hearing, bandwidth
Which of these two graphics are better for this product?	image	choice	visual
Which hairstyle makes you trust this politician more?	images	choice	visual
Do you think this commercial would make money?	video	profitable(0-1)	video, bandwidth
What age range do you think would like this song the most?	audio	age(5-90)	audio, bandwidth
Do you recognize any of these faces?	images	free tags, no scalar	visual

5. Conclusion and Implications

QuestionMarket has the potential to go beyond traditional web advertising with an increased level of profitability, create an online marketplace trading in units of knowledge answers, and create completely new ways to generate and aggregate demographic and opinion data.

5.1 Questions and Issues

There are many unanswered questions and potential issues with the QuestionMarket model:

- Do individual units of answers produce clean enough information as to be valuable?

- Is QuestionMarket easy enough to use for widespread use?

- Are people willing to spend more time answering questions in exchange for incentives? (Relating the need for incentives of average web survers to the self-organization techniques of knowledge-elites)

- Is there sufficient market demand - are there enough organizations needing questions answered?

- How is the compensation weighted to ensure that "click randomly" is not the dominant strategy?

- Will the current "fake" feedback banner ads poison the QuestionMarket market?

- How can we enable direct trading of services over a network?

    o In particular, what infrastructure would support a decentralized market in "human computation," harnessing people power to solve hard AI problems today?

    o Furthermore, could it also incentivize developers to plug in competitive automated services in the future?

Some of these questions can be approached through economics and game theory, while others may have to be field-tested during QuestionMarket's initial market testing.

5.2 Business Strategy

The business implications of QuestionMarket are two-fold: to leverage existing banner advertisement screen real-estate space to generate increased amount of revenue through answering questions, and to increase advertisement retention rates through asking questions about

the product being advertised. The initial beachhead of the QuestionMarket platform is direct competition against current methods of advertising, primarily interstitial advertising.

There are several business challenges, including:

- Identifying several markets that would benefit from low cost DHC. Potentially, these could be editors, obscenity filters, and anti-link spam efforts.

- For content providers, determining if the value provided from DHC is greater than the revenue form traditional advertising

- Connecting buyers to sellers in sufficient quantity as to reach critical mass for a sustainable marketplace.

5.2.1 Initial Market Segment

The early QuestionMarket economy is more likely to be successful if the early adopters are both information producers and consumers. Examples include content producers like the Wall Street Journal, where tagging story content would earn credits that are redeemable for access to premium content, or Google, where identifying linkspam might purchase a temporary KeyHole upgrade.

The markets where QuestionMarket has the best chance of a significant impact are those where there is a community with time that is valued low, but with demographic information that is valued highly. Specific examples include the music industry's continuing need for demographic information, especially the teen market. QuestionMarket will also attempt to leverage usage in situations where end-users may have some free time, indicating that mobile devices are a prime candidate for lightweight DHC interaction. To compliment the necessary low transaction cost of "farming out" units of DHC, similarly low transaction cost rewards should be used to motivate useful participation. Electronic rewards, such as credits towards ring tones, entry into sweepstakes, cell phone backgrounds, and other low-cost rewards would motivate quality response rates.

In addition to increased customer attentiveness to a product's ad, advertisers and marketers would also benefit from a targeted real-time source of demographic and micro-survey data. Current methods of calling individuals, organizing focus groups, and drawing on lists of opt-in members all have a variety of drawbacks, including time to set up, and cost limits on the number of responses. QuestionMarket could potentially provide a method of collecting tar-

geted demographic data that is as simple and immediate as choosing words on Google's Ad-Sense.

5.3   AI Research

A robust QuestionMarket economy buying and selling the answers to units of AI-Complete questions would have implications for online marketing, AI research, and the developing knowledge economy. The QuestionMarket marketplace becomes a continuous online tournament, benchmarking the "dollar value" of how good new AI algorithms are, an incentive that could provide AI research with a fertile testing ground. Successful AI research in AI-Complete problems would effectively collapse the new dimension of human computation for that specific class of problem, returning the problem to one of throwing enough computons at it. If a garage researcher thinks they have built a better obscenity filter, they can enter it into the QuestionMarket arena for a week, and see if at the end of the week the market values their entry's decisions enough as to be profitable.

Incremental Automation is also a beachhead for QuestionMarket. Questions that are currently not answerable by a computer, but require some specific expert knowledge, could be broken down into components. The first pass at answering a question could be done by a wide net of people, and the second pass done by dedicated resources, much like how seti@home re-checks any potential signal hits. Likewise, QuestionMarket could break into small chunks something like a difficult OCR problem, and have a dedicated resource reviewing the final re-sult. This has the potential to create outsourced filtering, a liquid worldwide market of a 'few seconds' of labor from the proverbial housewife in Bangalore that knows how to translate to a specific dialect. The potential benefit is symmetrical: AI checking human answers for quality, and/or humans checking to make sure the results from the AI are relevant.

6.   References and Notes

NASA Clickworkers

> "There are many scientific tasks that require human perception and common sense, but may not require a lot of scientific training. Identifying craters on Mars is something almost anyone can do, and classifying them by age is only a little harder. From November 2000 to September 2001, we ran an experiment that showed that public volunteers (clickworkers), many working for a few minutes here and there and others

choosing work longer, can do some routine science analysis that would normally be done by a scientist or graduate student working for months on end."

CMU's ESP Game and Peek-a-Boom

Labeling an image means associating word descriptions to it, as shown below. Computer programs can't yet determine the contents of arbitrary images, but the ESP game provides a novel method of labeling them: players get to have fun as they help us determine their contents. If the ESP game is played as much as other popular online games, we estimate that all the images on the Web can be labeled in a matter of weeks! Having proper labels associated to each image on the Internet would allow for very accurate image search, would improve the accessibility of the Web (by providing word descriptions of all images to visually impaired individuals), and would help users block inappropriate (e.g., pornographic) images from their computers.

Coase's Penguin, or, Linux and The Nature of the Firm (abstract), by Yochai Benkler in Yale Law Journal #112:

"Commons-based peer production, the emerging third model of production I describe here, relies on decentralized information gathering and exchange to reduce the uncertainty of participants. It has particular advantages as an information process for identifying and allocating human creativity available to work on information and cultural resources. It depends on very large aggregations of individuals independently scouring their information environment in search of opportunities to be creative in small or large increments. These individuals then self-identify for tasks and perform them for a variety of motivational reasons that I discuss at some length."

"I suggest that we are seeing is the broad and deep emergence of a new, third mode of production in the digitally networked environment. I call this mode **'commons-based peer-production,'** to distinguish it from the property- and contract-based models of firms and markets. Its central characteristic is that groups of individuals successfully collaborate on large-scale projects following a diverse cluster of motivational drives and social signals, rather than either market prices or managerial commands."

Scalability of human review

"Thanks to Adrian Holovaty for pointing out that Craigslist has this problem and has proven that a community review process is all you really need to keep bad things from happening."

Distributed Checksum Clearinghouse

"The DCC or Distributed Checksum Clearinghouse is an anti-spam content filter. As of mid-2004, it involves millions of users, tens of thousands of clients and more than 250

servers collecting and counting checksums related to more than 150 million mail messages on week days. The counts can be used by SMTP servers and mail user agents to detect and reject or filter spam or unsolicited bulk mail. DCC servers exchange or "flood" common checksums. The checksums include values that are constant across common variations in bulk messages, including 'personalizations.'"

PICS rating vocabulary

"A rating service is an individual, group, organization, or company that provides content labels for information on the Internet."

6.4 Acknowledgements

- CommerceNet for supporting the work.

- Dr. Rohit Khare for help with evolving the idea, making contacts, editing, and the education on P2P based markets.

- Dr. Jay "Marty" Tenenbaum for Lightweight/Incremental AI tie-ins to QuestionMarket

- Ryan Oprea for all the help with open market theory

- Tim Dennis and Steve Chan for the original PeerMarket work