

# An Introduction to Crowdsourcing for Language and Multimedia Technology Research

Gareth J. F. Jones

Centre for Next Generation Localisation  
School of Computing, Dublin City University, Dublin 9, Ireland  
gjjones@computing.dcu.ie,  
WWW home page: <http://www.computing.dcu.ie/~gjones>

**Abstract.** Language and multimedia technology research often relies on large manually constructed datasets for training or evaluation of algorithms and systems. Constructing these datasets is often expensive with significant challenges in terms of recruitment of personnel to carry out the work. Crowdsourcing methods using scalable pools of workers available on-demand offers a flexible means of rapid low-cost construction of many of these datasets to support existing research requirements and potentially promote new research initiatives that would otherwise not be possible.

**Keywords:** crowdsourcing, human computation, human language technologies, multimedia technologies

## 1 Introduction

Research in multimedia and language technologies often relies heavily on the use of datasets for training and evaluation of the various methods and algorithms proposed to achieve the objectives of the technology currently under consideration. For example, data for the training of machine translation or visual concept recognition algorithms or the evaluation of search effectiveness in information retrieval. The requirement for extensive manual involvement means that the development of these datasets is generally very expensive, and that the cost and logistics of the construction of such resources can represent a considerable obstacle to the exploration of new research directions which are not supported by existing training or test data resources.

Manual contributions to the construction of these datasets include for example the transcription of audio files as a first stage to them being used in the training of a speech recognition system, the labeling of visual concepts in images or frames of video for training of visual classifiers, and the writing of test queries for the evaluation of information retrieval algorithms for a specific document collection and judgement of relevance of the documents to each query. It is notable that these activities can generally be broken into small tasks which are often quite repetitive, and also that many of these tasks do not require any specific skills or rely only on an individual's existing skills. For example natural manual

translation of text by a bilingual speaker to build a training set for a machine translation system.

The nature of these tasks makes them highly suitable for online *crowdsourcing* methods where individual online workers are allocated and carry out small assignments in return for micro payments. Crowdsourcing is currently being used and explored in a number of areas of language and multimedia technology research. This work makes use of crowdsourcing for simple activities, e.g. speech transcription [20], but is also being explored for more challenging tasks, involving some level of personal creativity, e.g. writing search queries suitable for a particular document collection [6].

This paper provides an introduction to the general topic of crowdsourcing, highlights practical details of designing and using crowdsourcing activities, a short overview of some key existing work in crowdsourcing for language and multimedia research, and illustrates this with an example of the use of crowdsourcing in the MediaEval 2011 Rich Speech Retrieval task [15].

## 2 What is Crowdsourcing?

Crowdsourcing is form of *human computation*, where human computation is a method of having people do things that we might otherwise consider assigning to a computing device to calculate automatically, e.g. a language translation task. A *crowdsourcing system* facilitates a crowdsourcing process to complete a specified task. To carry out a task, a crowdsourcing system, enlists a “crowd” of human workers to help solve a defined problem [27]. Currently the best known crowdsourcing system is *Amazon Mechanical Turk (AMT)*<sup>1</sup>. For this reason, the crowdsourcing terminology of AMT is generally adopted in this paper.

In operation someone wishing to undertake a task defines the operations needed to complete it, and then divides them into multiple micro-tasks which when combined solve the problem at hand. The activity is then made available as a set of micro-tasks on a crowdsourcing system. Human workers are then recruited to undertake the micro-tasks for which they typically receive a payment for each completed task.

To operate successfully, a crowdsourcing system must address the following four issues [5]:

- How to recruit and retain workers?
- What contributions can the workers make?
- How to combine worker contributions to solve the target problems?
- How to evaluate workers and their contributions?

There are various forms of collaborative activities not all of which qualify as crowdsourcing. Collaborations between workers in crowdsourcing environments can be explicit or implicit. For example, in the development of Wikipedia or Linux, the crowdsourcing system enlists a crowd of workers to explicitly collaborate to build a long lasting artefact of use to a larger community [5]. By contrast,

<sup>1</sup> <https://www.mturk.com/>

workers in the ESP game implicitly collaborate to label images as a side effect while playing a game [37][36]. Similarly, workers using AMT collaborate implicitly, e.g. workers enlisted to find a missing boat in thousands of satellite images, where each worker inspects an individual image [5]. However, not all human-centric systems address these challenges, and such systems do not fall within the scope of crowdsourcing. For example, crowd management at a sports event does not look to recruit more members of the crowd, if anything in this case it would be preferable for members of the crowd to leave [5].

The availability of online crowdsourcing services such as AMT, is now making human computation resources available to various research communities. Among these communities there is currently significant interest in exploring the use of these services to support research activities in information and data processing technologies, and investigating how they might be used to open up new research directions, which might be technically innovative or may previously have been impractical using other means.

## 2.1 Creating and Managing a Crowdsourcing Activity

The basic process of creating a crowdsourcing activity is as follows. First, identify an activity which is amenable to being broken into small elemental tasks, e.g. the need to label the presence or absence of a visual concept in many thousands of images. Rather than give this elemental task to a specific individual, e.g. an employee, a crowdsourcing requester outsources it to someone else via the crowdsourcing system. The crowdsourcing system makes the tasks available to an ad hoc group of workers (who might be considered “employees” since they usually receive payment for their work) recruited via a call for participation. The workers bid for the offered work, the requester has the choice of which workers to accept to undertake the offered task. The decision of whether to accept an offer of work can be complex taking account of a number of factors; this issue is examined in more detail in Section 3. Once a worker has been accepted, they carry out the agreed work, the requester can then check the quality of the work; depending on the nature of the task, this checking process may itself be complex or non-exhaustive. Once the quality of the work has been checked, the requester then has the option to accept the work and make payment to the worker, or to reject it, in which case payment is not made. The decision as to whether to accept the work may not be straightforward, and is also discussed further below.

## 2.2 Why use Crowdsourcing?

Many areas of work involve a need to complete a large number of repetitious small tasks with high short term peak loads. In many cases these tasks do not involve specialist skills, with the key requirement being that the person undertaking them should be conscientious and seek to do the job to the best of their ability. Crowdsourcing is often ideal for this type of situation since it provides rapid access to a very flexible and cheap workforce enabling fast completion of tasks

at short notice without the need for the development of a long term employment infrastructure.

Established crowdsourcing platforms such as MTurk provide a framework which enables new tasks to be developed very rapidly using a standard set of interface components, and to be made available to workers straightaway with a well defined mechanism for making payments. This enables early stage experimentation to develop a crowdsourcing task for completion. For example, to develop a task for assessment of document relevance to some user information need. The initial version of the task can then be deployed on the crowdsourcing platform in a pilot study. The outputs of the workers attempts to complete this pilot task can then be examined, any problems or unexpected responses noted, the structure or content of the task iterated, and the task run again until the workers complete the task as required, at which point the full set of tasks can be offered for completion.

### **2.3 Who are the workers?**

A question that arises in respect of offering crowdsourcing tasks on an online platform such as MTurk where anyone can register as a potential worker and bid for work is, who are these workers? Surveys have found that initially they were frequently based in the USA, often based in the home for family reasons, and who probably undertook the work for the sake of interest, rather than for the financial reward. However, workers are increasingly diverse and international, where some of them are much more motivated by the available financial rewards. For example students based in countries with emerging economies [25][30].

## **3 High Level Issues in Crowdsourcing**

Effective use of crowdsourcing requires a number of high level issues to be addressed. Assuming that an activity amenable to crowdsourcing has been selected, and broken down into elemental tasks, then the requester of the work needs to consider the following points: the level of payment or other incentives for the work, the design of the interface and interaction design of the task, and the choice of crowdsourcing platform.

In addition to these very practical issues, the requester also needs to give careful consideration to management of human factors relating to workers, including: recruitment of workers, retention of good workers, quality control of the work, trust in and reliability of workers, and detection of poor quality work [35]. The remainder of this section examines each of these issues in more detail.

### **3.1 Recruitment**

As outlined earlier, once they are happy with the design of their task, the requester makes their task available to the registered workers on their selected crowdsourcing platform. Workers are able to search a list of currently available

tasks, and request to be assigned to a task that they want to do. The list of available tasks typically makes the registered identify of the requester available to the worker, as well as summary details of their history as a requester in terms of completed tasks, payment record to workers, etc. The requester is similarly shown summary details of a worker requesting their task, in terms of amount of successfully completed previous tasks, etc. Workers are thus able to determine whether a potential “employer” treats their workers fairly, and employers can see whether their potential employees are reliable high quality workers. Thus, requesters and workers with established strong reputations are likely to be successful in the recruitment process: the requester gets the worker they want, and the worker gets chosen for the task they want. A reputable requester can often have their pick from among reputable workers applying for their tasks.

Further to their general history of successful completion of tasks, other practical issues relating to the workers suitability for a specific task may need to be addressed in the recruitment process, e.g. for a language translation task, the worker must have the requisite level ability as a translator between the required languages. In order to ensure the workers skill level, the requester can set a qualifying test before agreeing to let the worker take the task, e.g. to translate and check some sample text. On other occasions, the quality of previous tasks completed can be taken as sufficient proof of the worker’s skills.

### 3.2 Reputation

As indicated in the previous section, workers and requesters can make selections based on each others reputations. Reputations within a crowdsourcing system are based on previous task activities which are made available at the selection stage. A worker’s reputation can act as an incentive for a requestor to accept their offer to undertake a task, and to trust the likely quality of their work. A requester can also select a worker based on their own previous experience of the worker.

In terms of the requestor’s reputation, as indicated above, the requester has the option of approving completed work or not, if they don’t approve the work, they don’t pay the worker. They may also have the option of paying a bonus to individual workers for completing work of exceptional quality or above some agreed standard as part of the task specification. The requester can approve the work, pay for it, and then not use, e.g. if it is clear that the worker expended reasonable effort in attempting to complete the task, but that the output is just not useable for some reason. The requester’s fairness in recognising the genuine effort of workers can have a long term effect on their success in recruiting for subsequent tasks. By contrast if they are perceived to reject large amounts of work, with or without good reason, then they can damage their reputation with workers, and workers may not request to take their tasks. This can create a dilemma for requesters, if they accept poor quality work then they can waste money, but if they set their standards too high, they may fail to attract enough workers to complete the full set of tasks required for their overall objective. One consequence of this situation is that it creates an incentive for requesters to

develop their task so that it is easy enough for reliable completion by suitably qualified workers, but still enables them to fulfill their overall objectives.

Overall, the reputation of both requesters and workers can be important for the success of a crowdsourcing ecosystem. A requester with a reputation for posting well structured and clearly described tasks, with fair and prompt payments is likely to prove popular with regular workers. A worker who undertakes tasks in a professional manner is likely to prove popular with requesters.

### 3.3 Payment and Incentives

As already stated workers generally undertake tasks for micro-payments - very small payments for completion of individual tasks. Workers may volunteer for an available task because it looks interesting, but often they will do so because it looks a good way to earn some money [30]. For the requester, offering a suitable level of payment is a trade-off between:

- Underpayment, for either or both of worker’s time or expertise.
- Overpayment, which may attract workers keen to earn money without undertaking the task properly.
- Sufficient payment to motivate the worker to complete the task conscientiously, with the possibility of offering bonus payments for excellent work.

Thus the requester needs to offer sufficient payment to attract enough suitably qualified workers, to motivate them to complete the task well, but not to make the task too attractive to undesirable workers who just want to make money without completing the work properly [35]. The relationship between incentives and quantity and quality of work is often complex. For example, it has been found that paying for completion of a complete instance of a task overall leads to more work being carried out than incremental payment for sub-tasks, and that greater payment may increase quantity, but not quality of work [21]. Further interesting work on this subject is described in [10].

This problem of workers volunteering for tasks with no intention to complete them properly, is a major concern in crowdsourcing, since not only does the work not get done properly or perhaps not at all, but depending on the nature of the task, it may be difficult for the requester to check this, and the requester wastes their money.

### 3.4 Detecting Poor Quality Work

As noted above, it is important that workers are actually capable of successfully completing their assigned task. While not all tasks require specific expertise or skills, it is important that the worker should undertake the task to the best of their ability. Workers may sign up for a task to earn money (and even pass a qualifying test if required), and then attempt to get paid without completing the task properly. Detecting so called “spam” work is an important issue in quality control. While some tasks require all work to be checked, for some other

tasks checking all work is impractical. For example, it is not possible to check all relevance assessments made in the development of an information retrieval test collection, even if it is practical to check the indicated relevant documents, it is not possible, without repeating the task, to check the reliability of the relevance judgements of all documents marked as non relevant.

One method to combat the problem of dishonest work is for requesters to set up “honey pots” with known answers. The worker completes these as part of their work, but unbeknownst to them, the requester knows the answer to the honey pot questions and can easily check for faked or poor quality work. The requester can then refuse to pay for the work, and bar the worker from undertaking further work for them. The worker’s reputation within the crowdsourcing system will also fall since their work has been rejected, making it more difficult for them to get work with other requesters. This is another reason why workers are less likely to request tasks from requesters with a poor payment record, since it affects the worker’s overall rating in the crowdsourcing ecosystems, making them less attractive to other requesters. Thus, not only do workers not want to work for requesters who don’t pay them if their work is not perfect, they also do not want to work for them because it can make it more difficult to get other work in the future. Honey pots can be used initially in a worker selection phase prior to fully engaging the worker to filter obvious cheating workers, but also continue to be used once the worker has been selected with the honey pots randomly distributed in with the main task assignments. This topic is examined in more detail in [38], and techniques for managing crowdsourcing workers and data quality are proposed in other studies including [38][34][28].

## 4 Crowdsourcing in Language and Multimedia Technology Research

A growing number of examples of the use of crowdsourcing methods have appeared in language and multimedia technology research in recent years. This section briefly reviews some of the most notable examples.

### 4.1 Language Technology Research

One of the most important early studies examining crowdsourcing in language technology research is described in [33], which demonstrated that non-expert workers can produce work of a similar quality to expert workers for natural language annotation tasks. Callison-Burch and Dredze [4] surveys contributions to the NAACL-2010 workshop on creating Speech and Language Data with Amazon’s Mechanical Turk, and highlights important factors which should be taken into account when designing effective crowdsourcing tasks.

A survey of research using crowdsourcing for speech research related tasks is contained in [26]. In more detailed studies of individual activities, the use of crowd workers to transcribe speech is explored in [20], while [7] examines the more challenging task of transcribing non-native read speech and spontaneous

speech. Further work on non-expert transcription is described in [23]. Focusing more on speed rather than accuracy, a method for real-time captioning of speech is described in [17]. The related task of collecting spoken resources using crowdsourcing methods is explored in [14].

The PodCastle system explores use of crowd-based correction of transcription errors and the use of these corrections to improve system training [8][24]. A more complex crowd-based retraining approach for a spoken language system is described in [22].

In information retrieval, relevance assessment for queries is a time consuming and human resource expensive activity which involves manual judgement of the relevance of a, potentially very large, number of documents to a user information need expressed in some form of search query. The nature of the relevance assessment tasks make crowdsourcing an attractive option to undertake human assessment of document relevance. This has been explored in a number of studies including [1]. The very large number of document viewed for short periods by crowdsource workers means that it is not possible to manually check all the assigned relevance labels, which means that issues of worker behaviour including motivation and reward are important, as for example examined in [9].

Crowdsourcing has also been studied in the context of evaluation of machine translation in [3], which showed a similar level of performance to gold standard judgements of translation quality. The training of statistical machine translation systems is reliant on the availability of parallel or at least comparable corpora in the languages for which the translation system is to be developed. The amount of such content suitable for the training of a statistical machine translation system is thus an important issue. This can present a significant problem for language pairs for which sufficient amounts of naturally occurring training data are not available. The use of crowdsourcing for the development of machine translation training data is presented in [13]. Another exploration of this topic incorporating an active learning method is described in [2]. Even when well trained, the output of machine translation systems is not ideal. In some applications manual correction of machine translation output forms part of a practical workflow. The use of crowdsourcing in correction of machine translation output is investigated in [18].

Another important area of language technology research is summarization. An investigation into use of crowdsourcing in summarization research is reported in [19], which concludes that while crowdsourcing was not effective for generating gold standard summaries for use in research, it is suggested that it is useful for studying patterns of human behaviour when creating summaries.

## 4.2 Multimedia Technology Research

In the area of multimedia research one of the consistently expensive activities is manual labelling of training and test data. Labelling of images using a game-based approach was explored in [37][36]. Further work in image annotation is reported in [35] and [28].

Image labelling using crowdsourcing can be noisy if workers are accidentally careless or deliberately do not attempt to do the work properly, or if correct labelling is simply difficult, for example, if the image is unclear or assignment of the label uncertain for some other reason. Methods to improve label quality are proposed in [31][11][29].

The use of crowdsourcing to improve, extend and share automatically detected concepts in video fragments is examined in [32]. A crowdsourced human validation of image search results is described in [41]. Exploring the area of summarization in video access, [40] describes a method for rapid generation of video summaries incorporating the viewer’s preference.

In the emerging area of affect in multimedia, [34] describes a novel method of affective annotation of video using crowdsourcing. An initial investigation into the topic of crowdsourcing for user studies is reported in [12].

Beyond these subjects, other interesting topics to which crowdsourcing is being applied include social data analysis [39].

## 5 Crowdsourcing Platforms

Although Amazon Mechanical Turk (AMT) is generally the most popular crowdsourcing platform, a number of others are available which offer crowdsourcing functionality; these include: Crowdflower<sup>2</sup>, CloudCrowd<sup>3</sup>, DoMyStuff<sup>4</sup>, Clickworker<sup>5</sup>, Smartsheet<sup>6</sup>, uTest<sup>7</sup>, Elance<sup>8</sup>, oDesk<sup>9</sup>, and Freelancer<sup>10</sup>. Many of these are more concerned with flexible commercial recruitment services which appear less useful for scientific research. At present, the main alternative to AMT for the type of research examined in this paper is CrowdFlower.

A good way to get started with crowdsourcing is sign up as a worker on one of these platforms and do some tasks, and to understand the practical issues which arise by monitoring discussion forums.

While other services are available, the remainder of this paper focuses on the use of AMT, however the principles are general and can be applied to any similar crowdsourcing platform. AMT has been online since 2005 with an on-demand, scalable, and real-time workforce. It can be accessed via a “dashboard” GUI or using a programmers’ API. A requestor wanting to recruit workers to undertake a task creates a Human Intelligence Task (HIT), which is a web form composed of a number of instructions. AMT HITs are undertaken by workers referred to for AMT as “turkers.” The requestor specifies the reward which is available for completing the HIT.

<sup>2</sup> <http://crowdflower.com/>

<sup>3</sup> <http://www.cloudcrowd.com/>

<sup>4</sup> <http://www.domystuff.com/>

<sup>5</sup> <http://www.clickworker.com/en/>

<sup>6</sup> <http://www.smartsheet.com/>

<sup>7</sup> <http://www.utest.com/>

<sup>8</sup> <https://www.elance.com/>

<sup>9</sup> <https://www.odesk.com/>

<sup>10</sup> <http://www.freelancer.com/>

## 6 Implementation of an AMT HIT

This section outlines a suggested procedure for the development and usage of an effective AMT HIT. This is based partially on principles of user-centered design for interaction systems and practical experience working with AMT HITs.

When building a completely new HIT, it can be useful to build a mock up which could be as simple as a list of the instructions to be completed by the worker, and test it locally with your research team. Feedback from these tests can then be incorporated to refine the task. Subsequently a limited test run can be implemented on AMT with a very small dataset. Key questions to be considered include:

- What is the time for completion of the HIT? If replicated by other workers, how long would it take to complete the overall task with a set of HITs?
- Do people understand the task?
- Consider needs for quality control:
  - Is a qualification test needed?
  - Adjust qualification passing grade and/or acceptable approval rate of workers if necessary.
- Check suitability or correctness of the output.
- Look for spammers.
  - Are gold answers (honey pots) needed to catch spam workers?
- Look at comments from workers: are they happy / unhappy?, would they work for you again?, is the HIT too easy / too hard?
- Is the payment rate for work - too high / too low? Do people sign up or not? Do they complain after doing the task that the payment is too small? Is there evidence of poor workers attracted by apparently high payment for the work?
- Address feedback, e.g. poor guidelines, payments, passing grade, etc. - email exchange. Everything counts! The HIT is only as good as the weakest part!
- Run another experiment with new settings and the same data to make sure that everything is now working as expected, if not, iterate again.
- If all is in order, launch a full batch of the HIT.

There are many tasks active on AMT with corresponding HITs on offer at any point. If you want to attract the best workers, you need to grab their attention! Make the HIT look attractive in some way, make it sound interesting, but be honest, workers accepting and completing HITs which are not accurately represented in the description are prone to post complaints, which may affect your ability as a requester to attract more workers in the future.

Split a large crowdsourcing activity into batches of HITs; only have one batch in the system at a time. There are only a limited number of workers available at any time, having multiple batches of the same HIT available in parallel is unlikely to complete the overall activity any faster since it will divide the available effort. Also, running batches of HITs sequentially means that you can review feedback from batch  $n$  before running batch  $n+1$ , allowing you to make small adjustments to the HIT if needed.

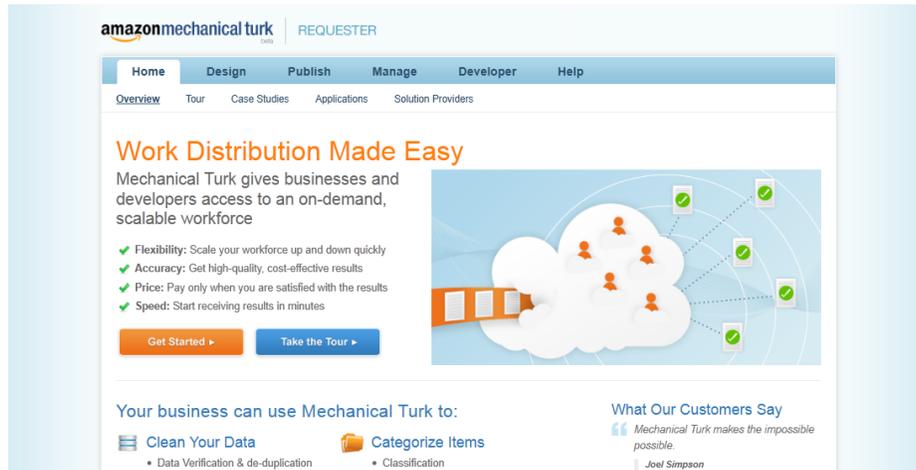


Fig. 1. AMT Requester Entry Point.

## 7 Crowdsourcing Example

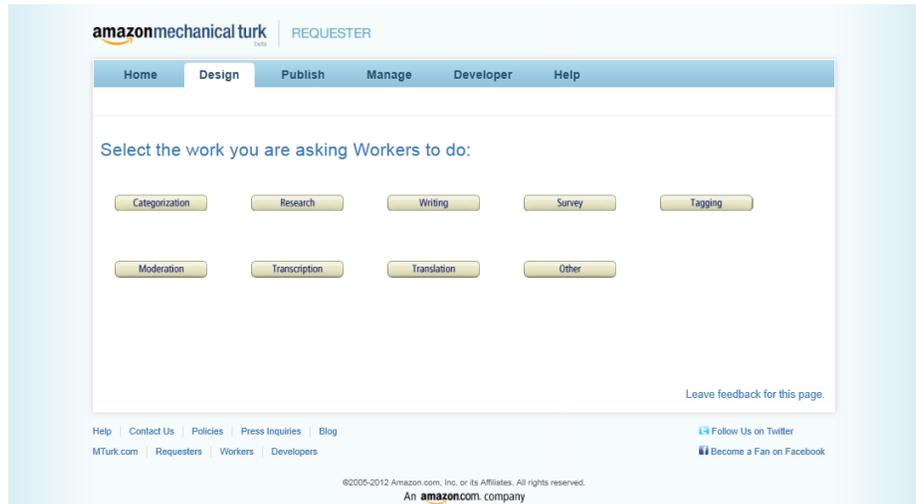
This section describes the development and use of an AMT HIT for collection of a query set for the MediaEval 2011 Rich Speech Retrieval (RSR) task [15]. RSR was offered as a benchmark task as part of the MediaEval multimedia evaluation benchmark<sup>11</sup>. Registered task participants were required to try to identify a single video known to be relevant to a searcher’s information need in a known-item search task, and to locate the optimal point to start playback within the video, referred to as a jump-in point. This task models a user trying to re-find a previously viewed segment of video. The RSR task wished to explore five different functions of speech, represented as illocutionary speech acts: ‘apology’, ‘definition’, ‘opinion’, ‘promise’ and ‘warning’. The document set consisted of 1974 episodes (247 dev, 1727 test) 350 hours of semi-professional video harvested from blip.tv [16], available for download under a creative commons licence.

An AMT HIT was used to develop the RSR test collection by locating a number of interesting jump-in points in the video collection, describing them and forming a search topic statement for each one. The evaluation task was then to use the topic statement to try to locate the jump-in point. This section gives more details of the task and the AMT HIT used to develop the test collection.

### 7.1 Setting up an AMT HIT

The first step is for the requester to enter the AMT system, as shown in Figure 1. Once in the system, the requester selects a category of task which they wish to undertake, see Figure 2.

<sup>11</sup> <http://www.multimediaeval.org>



**Fig. 2.** AMT Options for Design of a HIT.

An AMT HIT is constructed on an HTML form. Several options are available to the requester to develop their HIT:

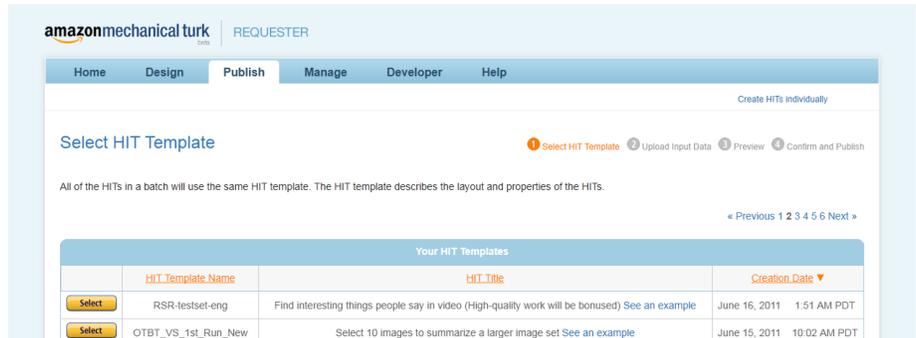
- Use one of the templates provided within AMT.
- Download one of the templates from AMT, edit it and upload the edited HIT to AMT.
- Write their own, and upload it to AMT.
- Select an existing template that they uploaded previously.

Figure 3 shows the HITs designs available to the requester.

The MediaEval 2011 RSR HIT was written from scratch and uploaded to AMT. Once uploaded the requester can preview the HIT as it will be seen by the worker, see Figure 4. When the requester is satisfied with the HIT, they can publish it to make it available to the workers, see Figure 5. Note that for AMT, the requester must have sufficient credit registered in the system to pay for completion of the batch before it can be made available to workers.

Once a HIT has been selected and made available to workers, the requester can monitor the progress of the current batch of the instance of the HIT. Figure 6 shows the percentage of the requested task completed so far by the workers. This also shows the average time to complete a HIT and the average rate of pay for the work. This information is useful to the requester in assessing the effectiveness of the HIT and to potential workers to decide whether to apply for the HIT.

Completion of the overall RSR query construction task required multiple workers to complete the HIT by viewing different video files. Operation of the HIT required input variables to specify the video to be viewed in this instance of the HIT. The variables gave details of the path to the server where the video



**Fig. 3.** AMT List of available HIT designs.

to be viewed in the HIT was stored. The code also specified the video player to be used and the video to be played.

A set of HITs to be completed in a single batch are defined in a csv file uploaded by the requester. The values of the variables (the names of the videos in this case) are specified in the file. The csv file to be used in this batch of the HIT is selected using the interface shown in Figure 9. During execution of the batch AMT keeps track of which HITs have been completed. Workers are assigned HITs until either the batch has been completed, or the requester stops execution of further HITs in this batch.

## 7.2 Details of the MediaEval 2011 RSR HIT

This section gives details of the MediaEval 2011 RSR HIT. The version of the HIT shown here was used for collection of the experimental dataset, and was developed iteratively using batch trial runs. The development and use of the HIT had a research element to it. The examples of crowdsourcing in language and multimedia technology research outlined earlier are generally simple tasks where workers are asked to objectively label linguistic or visual content. By contrast, the RSR HIT was exploring the behaviour and effectiveness of untrained crowdsource workers in a creative process, in this case the generation of the search queries to look for the known-item. The ability of carry out research activities requiring creative input potentially greatly increases the scope for use of crowdsourcing in research.

**MediaEval 2011 RSR HIT**

Find interesting things people say in videos.

Imagine that you are watching videos on YouTube, when you come across something interesting you might want to share it on Facebook, Twitter or your favourite social network.

Now please watch this video and search for an interesting video segment that you would like to share with others because it is:

an apology, full example  
 a definition, full example  
 an opinion, full example  
 a promise, full example  
 a warning, full example

(you can move your mouse over the words for text-only examples and click for full example with video)

The selected segment should be around 10-30 seconds long. Don't be alarmed if the video doesn't start at the beginning (and also don't scroll back).

When you are finished with answering the questions, don't forget to click the "Submit" button at the bottom of the page. Thank you very much for your help!

1) What kind of segment is the video part that you selected?

- an apology
- a definition
- an opinion
- a promise
- a warning
- I can't find anything like this in this video

2) We can improve our task by excluding this video. Only if you chose "I can't find anything like this in this video", please give us a reason why and tell us if you think other people will have the same problem (one or two sentences, please be as neutral as possible in your description), and you should skip the follow-up questions.

3) For your selected segment (in 1) above, what is the start time (please specify exactly in minutes and seconds)? Please pay attention to the time shown in the left corner of the bottom line of the video player.

Minute \_  
 Second \_

4) For your selected segment (in 1) above, what is the end time (please specify exactly in minutes and seconds)? Please pay attention to the time shown in the left corner of the bottom line of the video player.

Minute \_  
 Second \_

5) What was said during your selected segment? Please write down the exact words the speaker is saying (please transcribe precisely). If you are not sure what the exact word was, please write down what you think the word was and mark it with a star (for example, 'French president \*Sarkosie was saying ...' if you are not sure how to spell the name 'Sarkozy' properly)

6) When sharing this particular part of the video (your selected segment) on a social network, what comment would you add to the video to make sure that your friends have an idea what the video segment is about?

Please do not use informal internet language (such as '4 u' instead of 'for you').

Be as objective as possible when describing the video segment and do not express your personal opinion/attitude, either positive or negative.

7) Imagine you would like to search for similar video segments using a search engine (such as Google, Bing, Yahoo) what would you put in the search box?

We understand that this work requires a lot of your time and concentration, so we would like to bonus the high-quality of your results.

Please tell us your opinion about the size of bonus you deserve. Choose and justify your choice. Please keep in mind that we are carrying out non-profit university research (we can afford a maximum of 21 cents bonus, but only for really excellent responses).

When making our decision on your bonus level we create a compromise between our budget and your request.

- 0 cents
- 7 cents
- 11 cents
- 21 cents

### 7.3 Details of Completed HITs

The requester can view a summary of the workers who have selected and completed the HIT and their activities, shown in Figure 8. A more detailed version can be downloaded in a csv file.

The requester can review the HIT exactly as seen by the worker, the next slide, shown Figure 9. In the case of the RSR HIT, this means that the requester can see the specific video viewed by the worker. This enables the requester to check and confirm the details of the work delivered.

When deciding whether to select a HIT, a worker can view the currently available HITs, shown in Figure 10. This shows:

- a brief summary of the HIT,
- whether there is a qualification requirement,
- the time allotted for completing it,
- the available reward for the HIT,
- no of HITs available for the batch,

- and for partially completed batches, the payment per hour made to workers so far.

Requesters may also provide potential workers with a sample HIT so that they know what they will need to do.

#### 7.4 Notes on the MediaEval 2011 RSR HIT

There was no qualification requirement for the HIT. However, workers had to have a 90% acceptance rate by requesters for their previous work. A small scale initial run received negative feedback because the reward payment was judged to be too low for the work required. Increasing the reward for a subsequent batch of HITs fixed this problem. Also, indicating that the work was being carried out for a non profit organisation (a university) meant that workers were more willing to accept the level of payment on offer.

Workers were allowed to select their own bonus from several available to reflect the quality of their work. Workers were generally found to be honest and good judges of the bonus their work deserved, interestingly many did not request the maximum available bonus payment.

Some problems were encountered due to the need to play an external video. these included:

- issues with the worker’s browser.
- issues with their equipment, e.g. audio playback.
- issues with bandwidth required to play the video.

Workers were allowed to indicate if they were unable to find one of the target speech acts in the video they were given: In which case they did not need to complete the HIT form. Subsequent checking showed that they were generally correct in their judgement. However, some spamming of the HIT was found. In these cases workers were clearly not attempting to complete the HIT properly, and they were not paid in these cases. In some cases workers had completed the HIT, but their work was not found to be useable in the RSR task. In these cases, since they had clearly honestly attempted to complete the task, payment was made. Overall the use of this HIT for development of the MediaEval 2011 RSR was found to be successful, with a general conclusion being that AMT workers are able to undertake carefully designed creative tasks. A more detailed description of the design of the MediaEval 2011 RSR test collection is contained in [6].

## 8 Resources Available to Support Use of Crowdsourcing

A large number of resources are available from the Crowdsourcing News, Events, and Resources website maintained by Matt Lease at: <http://ir.ischool.utexas.edu/crowd>. This includes slides from conference and workshop tutorials and keynotes, lists of recommended readings, etc. The CrowdScope wiki providing links useful links to crowdsourcing resources is available at <http://crowdscope.org/>.

## Acknowledgements

Some of the background material on crowdsourcing for information retrieval draws on presentations by Matt Lease and Omar Alonso, and is used with permission. Thanks to Matt Lease and Martha Larson for invaluable discussions on the topic of crowdsourcing. Many thanks to Maria Eskevich for providing the screen shots for the MediaEval 2011 Rich Speech Retrieval task crowdsourcing example. This work was supported by Science Foundation Ireland (Grant 08/RFP/CMS1677) Research Frontiers Programme 2008 and (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (CNGL) project at DCU.

## References

1. O. Alonso, D. E. Rose, and B. Stewart. Crowdsourcing for relevance evaluation. *SIGIR Forum*, 42(2):9–15, 2008.
2. V. Ambati, S. Vogel, and J. Carbonell. Active learning and crowd-sourcing for machine translation. In *Proceedings of the Seventh international conference on Language Resources and Evaluation (LREC 2012)*, pages 2169–2174, 2010.
3. C. Callison-Burch. Fast, cheap, and creative: Evaluating translation quality using Amazon’s Mechanical Turk. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP 2009)*, pages 286–295, 2009.
4. C. Callison-Burch and M. Dredze. Creating speech and language data with Amazon’s Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk (CSLDAMT 2010)*, pages 1–12, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
5. A. Doan, R. Ramakrishnan, and A. Y. Halevy. Crowdsourcing systems on the World-Wide Web. *Communications of the ACM*, 54(4):86–96, 2011.
6. M. Eskevich, G. J. F. Jones, M. Larson, and R. Ordelman. Creating a Data Collection for Evaluating Rich Speech Retrieval. In *Proceedings of the Eighth international conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey, 2012.
7. K. Evanini, D. Higgins, and K. Zechner. Using Amazon Mechanical Turk for transcription of non-native speech. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk (CSLDAMT 2010)*, pages 53–56. Association for Computational Linguistics, 2010.
8. M. Goto and J. Ogata. Podcastle: Recent advances of a spoken document retrieval service improved by anonymous user contributions. In *Proceedings of Interspeech 2011*, 2011.
9. C. Grady and M. Lease. Crowdsourcing document relevance assessment with Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk (CSLDAMT 2010)*, pages 172–179. Association for Computational Linguistics, 2010.
10. J. J. Horton. Employer expectations, peer effects and productivity: Evidence from a series of field experiments. *CoRR*, abs/1008.2437, 2010.
11. P. G. Ipeirotis, F. Provost, V. Sheng, and J. Wang. Repeated Labeling Using Multiple Noisy Labelers. *SSRN eLibrary*, 2010.

12. A. Kittur, E. H. Chi, and B. Suh. Crowdsourcing user studies with Mechanical Turk. In *Proceedings of the 2008 Conference on Human Factors in Computing Systems (CHI 2008)*, pages 453–456. ACM, 2008.
13. A. Kunchukuttan, S. Roy, P. Patel, K. Ladha, S. Gupta, M. M. Khapra, and P. Bhattacharyya. Experiences in resource generation for machine translation through crowdsourcing. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, Istanbul, Turkey, 2012.
14. I. Lane, A. Waibel, M. Eck, and K. Rottmann. Tools for collecting speech corpora via Mechanical-Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk (CSLDAMT 2010)*, pages 184–187. Association for Computational Linguistics, 2010.
15. M. Larson, M. Eskevich, R. Ordelman, C. Kofler, S. Schmiedeke, and G. J. F. Jones. Overview of Mediaeval 2011 Rich Speech Retrieval Task and Genre Tagging Task. In *MediaEval 2011 Workshop Notes Proceedings*, volume 807. CEUR-WS.org, 2011.
16. M. Larson, M. Soleymani, M. Eskevich, P. Serdyukov, R. Ordelman, and G. J. F. Jones. The Community and the Crowd: Multimedia Benchmark Dataset Development. *IEEE Multimedia*, 19(3):15–23, July 2012.
17. W. S. Lasecki, C. D. Miller, A. Sadilek, A. Abumoussa, D. Borrello, R. Kushalnagar, and J. P. Bigham. Real-time captioning by groups of non-experts. In *Proceedings of 25th ACM Symposium on User Interface Software and Technology (UIST’12)*, pages 23–34, Cambridge, Massachusetts, USA, 2012. ACM.
18. S. Liao, C. Wu, and J. M. Huerta. Evaluating human correction quality for machine translation from crowdsourcing. In *Recent Advances in Natural Language Processing (RANLP 2011)*, pages 598–603, 2011.
19. E. Lloret, L. Plaza, and A. Aker. Analyzing the capabilities of crowdsourcing services for text summarization. *Language Resources and Evaluation (LRE)*, 2012.
20. M. Marge, S. Banerjee, and A. I. Rudnicky. Using the Amazon Mechanical Turk for transcription of spoken language. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2010)*, pages 5270–5273. IEEE, 2010.
21. W. Mason and D. J. Watts. Financial incentives and the “performance of crowds”. In *Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP ’09*, pages 77–85, New York, NY, USA, 2009. ACM.
22. I. McGraw, S. C. an Panupong Pasupat, J. Liu, and J. Glass. Automating crowd-supervised learning for spoken language systems. In *Proceedings of Interspeech 2012*, 2012.
23. S. Novotney and C. Callison-Burch. Cheap, fast and good enough: automatic speech recognition with non-expert transcription. In *The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT 2010)*, pages 207–215, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
24. J. Ogata and M. Goto. PodCastle: Collaborative training of language models on the basis of the wisdom of crowds. In *Proceedings of Interspeech 2012*, 2012.
25. G. Paolacci, J. Chandler, and P. G. Ipeirotis. Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5):411–419, 2010.
26. G. Parent and M. Eskenazi. Speaking to the Crowd: looking at past achievements in using crowdsourcing for speech and predicting future challenges. In *Proceedings of Interspeech 2011*, pages 3037–3040, 2011.
27. G. Pickard, W. Pan, I. Rahwan, M. Cebrian, R. Crane, A. Madan, and A. Pentland. Time-critical social mobilization. *Science*, 334(6055):509–512, October 2011.

28. C. Rashtchian, P. Young, M. Hodosh, and J. Hockenmaier. Collecting image annotations using Amazon’s Mechanical Turk. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon’s Mechanical Turk*, CSLDAMT ’10, pages 139–147, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
29. V. C. Rayker, S. Yu, L. H. Zhao, G. Hermosillo Valadez, C. Floring, L. Bogoni, and L. May. Learning from crowds. *Journal of Machine Learning Research*, 11:1297–1322, April 2010.
30. J. Ross, L. Irani, M. S. Silberman, A. Zaldivar, and B. Tomlinson. Who are the crowdworkers?: shifting demographics in Mechanical Turk. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, CHI EA ’10, pages 2863–2872, New York, NY, USA, 2010. ACM.
31. V. S. Sheng, F. Provost, and P. G. Ipeirotis. Get another label? improving data quality and data mining using multiple, noisy labelers. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD ’08, pages 614–622, New York, NY, USA, 2008. ACM.
32. C. G. Snoek, B. Freiburg, J. Oomen, and R. Ordelman. Crowdsourcing rock n’ roll multimedia retrieval. In *Proceedings of the 18th ACM International Conference on Multimedia (ACM MM 2010)*, pages 1535–1538. ACM, 2010.
33. R. Snow, B. O’Connor, D. Jurafsky, and A. Y. Ng. Cheap and fast — but is it good?: Evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2008)*, pages 254–263, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.
34. M. Soleymani and M. Larson. Crowdsourcing for Affective Annotation of Video: Development of a Viewer-reported Boredom Corpus. In V. Carvalho, M. Lease, and Yilmaz, editors, *Proceedings of the SIGIR 2010 Workshop on Crowdsourcing for Search Evaluation (CSE 2010)*. ACM, 2010.
35. A. Sorokin and D. Forsyth. Utility data annotation with Amazon Mechanical Turk. In *Proceedings of the First IEEE Workshop on Internet Vision at CVPR 08*, pages 1–8. IEEE, 2008.
36. L. von Ahn. Games with a Purpose. *Computer*, 39(6):92–94, June 2006.
37. L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI ’04, pages 319–326, New York, NY, USA, 2004. ACM.
38. J. Wang, P. G. Ipeirotis, and F. Provost. Managing crowdsourcing workers. In *Proceedings of the Winter Conference on Business Intelligence*, 2011.
39. W. Willett, J. Heer, and M. Agrawala. Strategies for crowdsourcing social data analysis. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI ’12, pages 227–236, New York, NY, USA, 2012. ACM.
40. S.-Y. Wu, R. Thawonmas, and K.-T. Chen. Video summarization via crowdsourcing. In *Extended Abstracts on Human Factors in Computing Systems (CHI 2011)*, pages 1531–1536. ACM, 2011.
41. T. Yan, V. Kumar, and D. Ganesan. Crowdsearch: exploiting crowds for accurate real-time image search on mobile phones. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, MobiSys ’10, pages 77–90, New York, NY, USA, 2010. ACM.

Home Design **Publish** Manage Developer Help

Create HITS individually

## Preview HITS

1 Select HIT Template 2 Upload Input Data **3 Preview** 4 Confirm and Publish

This is how your HIT will look to workers. Make sure that any variables in the HIT are correctly replaced by your input data, then click "Next".

**RSR-testset-eng**

Find interesting things people say in video (High-quality work will be bonused)

Requester: [REDACTED] Reward: \$0.19 per HIT HITS available: 3278 Duration: 40 Minutes

Qualifications Required: HIT Approval Rate (%) for all Requesters' HITS greater than or equal to 90

### HIT Preview

**Find interesting things people say in videos**

Imagine that you are watching videos on YouTube. When you come across something interesting you might want to share on Facebook, Twitter or your favorite social network. Now please watch this video and search for an interesting video segment that you would like to share with others because it is:

- [an apology, full example](#)
- [a definition, full example](#)
- [an opinion, full example](#)
- [a promise, full example](#)
- [a warning, full example](#)

(you can move your mouse over the words for text-only examples and click for full example with video)  
 The selected segment should be around 10-30 seconds long.  
 Don't be alarmed if the video doesn't start at the beginning (and also don't scroll back).  
 When you are finished with answering the questions, don't forget to click the "Submit" button at the bottom of the page. Thank you very much for your help!

1) What kind of segment is the video part that you selected?

an apology  a definition  an opinion  a promise  a threat  I c this vi

2) We can improve our task by excluding this video. **Only** if you choose in this video, please give us a reason why and tell us if you think other problem (one or two sentences, please be as neutral as possible in your skip the follow-up questions.

3) For your selected segment, what is the start time (please specify in seconds)? Please pay attention to the time shown in the left corner of the video player.

Showing HIT 1 of 3278

Fig. 4. Previewing the HIT.

amazonmechanicalturk lets REQUESTER

Home Design **Publish** Manage Developer Help Create HITs individually

Your account does not have sufficient funds to publish these HITs. [Click here to fund your account](#)

### Confirm and Publish Batch

1 Select HIT Template 2 Upload Input Data 3 Preview 4 **Confirm and Publish**

Please review the information about the HIT batch, then click "Publish HITs".

**RSR-testset-eng**

Batch Summary

Batch Name:  Description:

**Batch Properties**

Title: Find interesting things people say in video (High-quality work will be bonused)

Description: We are carrying out non-profit research at a university about the things that people say in videos that are worth sharing. The HIT contains a video served over http: your browser (avoid Safari) must be set to display http within https.

Batch expires in: 14 Days

Results are automatically approved after: 14 Days

Workers must meet the following Qualifications to work on these HITs: HIT Approval Rate (%) for all Requesters' HITs score greater than or equal to 90

**HITs**

Number of HITs in this batch:	3,278
Number of assignments per HIT:	x 1
Total number of assignments in this batch:	3,278

**Cost**

Reward per Assignment:	\$0.190
	x 3,278 <small>(total number of assignments in this batch)</small>
Estimated Total Reward:	\$622.820
Estimated Fees to Mechanical Turk:	+ \$62.282 <small>(fees paid to Mechanical Turk) (see details)</small>
Estimated Total Cost:	<b>\$685.102</b> <small>(this is the amount that will be deducted from your Available Balance when you click "Publish HITs")</small>
Your Available Balance:	\$14,684 <small>(before clicking "Publish HITs")</small>
Your Projected Balance:	\$-670.418 <small>(after clicking "Publish HITs")</small>

[Help](#) | [Contact Us](#) | [Policies](#) | [Press Inquiries](#) | [Blog](#)  
[MTurk.com](#) | [Requesters](#) | [Workers](#) | [Developers](#)

[Follow Us on Twitter](#)  
[Become a Fan on Facebook](#)

Timed Out: ©2009-2012 Amazon.com, Inc. or its Affiliates. All rights reserved.

**Fig. 5.** Publishing HIT to be available to Workers.

'RSR-testset-eng'

View the latest status of this batch, make changes, or get results.

We are carrying out non-profit research at a university about the things that people say in videos that are worth sharing. The HIT contains a video served over http; your browser (avoid Safari) must be set to display http within https.

**Status**

Status: Pending  
Review: 14% submitted, 100% published

Assignments Completed: 476 / 3,382  
Creation Time: June 16, 2011 1:55 AM PDT  
Average Time per Assignment: 10 minutes 12 seconds  
Completion Time: July 03, 2011 7:45 AM PDT  
Average Hourly Rate: \$1.118 (Cancelled)

**Settings**

**RSR-testset-eng**

Description: We are carrying out non-profit research at a university about the things that people say in videos that are worth sharing. The HIT contains a video served over http; your browser (avoid Safari) must be set to display http within https.

Keywords: video, quotations, PetaMedia

Qualification Requirement: HIT Approval Rate (%) for all Requesters' HITs greater than or equal to 90

Number of Assignments per HIT: 1  
Reward per Assignment: \$0.190  
Input File: 2011-06-13\_rsr\_\_testset-eng-only.csv

HIT expires on: EXPIRED  
Assignment duration: 40 minutes  
Auto Approval Delay: 14 days

**Results**

Assignments pending review: 25  
Assignments approved: 252  
Assignments rejected: 199

**Cost Summary**

Estimated Total Reward: \$622.820  
Estimated Fees to Mechanical Turk: \$62.282 (see details)  
Estimated Total Cost: \$685.102  
These costs are only an estimate until all of the assignments have been submitted and reviewed.

Find: RSR | [Next](#) | [Previous](#) | [Highlight all](#) | [Match case](#)

Fig. 6. AMT progress of the current batch of the HIT.

amazonmechanicalturk.com REQUESTER

Home Design Publish Manage Developer Help

Create HITs individually

**Upload Input Data**

1 Select HIT Template 2 Upload Input Data 3 Preview 4 Confirm and Publish

Choose a .csv file with the variables you specified in your HIT Template ("RSR-testset-eng") (learn more).

Your file's column headers need to include: video, starttime, duration

**RSR-testset-eng**

Locate a New File

Download a sample of the input file for this HIT template

« Previous 1 2 3 4 5 6 Next »

Or Choose an Existing File

Your Existing Files					
	File Name	Input Lines	Creation Date		
<input type="button" value="Select"/>	otbt_image_variables_loc_161_207.csv	47	August 12, 2011 3:49 AM PDT	<input type="button" value="Delete"/>	
<input type="button" value="Select"/>	otbt_image_variables_loc_121_160.csv	40	August 09, 2011 5:15 AM PDT	<input type="button" value="Delete"/>	
<input type="button" value="Select"/>	otbt_image_variables_loc_101_120.csv	20	August 03, 2011 2:42 PM PDT	<input type="button" value="Delete"/>	

Fig. 7. Selecting a csv file for a batch of HIT runs.

**Review Results**

Select the check boxes on the left to approve or reject results. You only pay for approved results. To evaluate results offline, select Download CSV.  
For additional batch information, [view batch details](#).

'RSR-testset-eng'

Customize View Filter Results Upload CSV Approve All Download CSV

25 of 476 assignments (FILTER APPLIED: only show assignments that are in "Submitted" status)

<input type="checkbox"/>	HIT ID ▲	Worker ID	Lifetime Approval Rate	Input Video	Input Duration	Input Length	Input Startmin	Input Sta
<input type="checkbox"/>	21405BMJUH1Y16PL7IADS5WAZ2QLW	A956X2WUE0607	100% (2/2)	Lauraflanders-TheFWordSeptember92008677.flv.ogv	160	7	0	0
<input type="checkbox"/>	214N9U8P9Y38TVMTE1U4IM07Z2FFWO	AV2OS09PKWD68	100% (7/7)	ReporterRoundtable-RREps225184.flv.ogv	1652	7	14	840
<input type="checkbox"/>	27A1CSTM0FXRDRUE825YJNKJ3STPSM	A13B0ND5GVEG7K	100% (1/1)	Robblatt-PodcastingsDeadAPodCampMontrealPrevie...	92	7	0	0

Find: RSR | Next Previous Highlight all Match case

Fig. 8. Requester Results.

**Review Results**

Select the check boxes on the left to approve or reject results. You only pay for approved results. To evaluate results offline, select Download CSV.  
For additional batch information, [view batch details](#).

'RSR-testset-eng'

Customize View Filter Results Upload CSV Approve All Download CSV

25 of 476 assignments (FILTER APPLIED: only show assignments that are in "Submitted" status)

**Find interesting things people say in videos**

Imagine that you are watching videos on YouTube. When you come across something interesting you might want to share on Facebook, Twitter or your favorite social network. Now please watch this video and search for an interesting video segment that you would like to share with others because it is:

- [an apology, full example](#)
- [a definition, full example](#)
- [an opinion, full example](#)
- [a promise, full example](#)
- [a warning, full example](#)

(You can move your mouse over the words for test-only examples and click for full example with video)  
The selected segment should be around 10-30 seconds long.  
Don't be alarmed if the video doesn't start at the beginning (and also don't scroll back).  
When you are finished with answering the questions, don't forget to click the "Submit" button at the bottom of the page. Thank you very much for your help!

1) What kind of segment is this?

an apology  a definition

Find: RSR | Next Previous Highlight all Match case

Fig. 9. Requester's view of the HIT as seen by the Worker.

The screenshot displays the Amazon Mechanical Turk interface. At the top, there are navigation tabs for 'Your Account', 'HITS', and 'Qualifications'. A notification indicates '159,693 HITS available now'. Below the navigation is a search bar with the text 'Find HITS containing' and a filter for 'that pay at least \$ 0.00'. The main section is titled 'All HITS' and shows '1-10 of 1906 Results'. The results are sorted by 'HITS Available (most first)'. The table below lists five HITs with their respective details:

HIT Title	Requester	HIT Expiration Date	Time Allotted	Reward	HITs Available
3 questions about your city UNDER 230,000 population only = \$0.17 bonus*** - qualification instantly granted (no wait)	WSDVC.COM	Jan 30, 2012 (6 days 18 hours)	3 hours	\$0.00	20533
Inv 8_2	rgbzkt0d	Feb 8, 2012 (2 weeks 2 days)	48 minutes	\$0.00	17032
Classify Webpages [Please read description]	Vijay Krishnan	Feb 2, 2012 (1 week 3 days)	60 minutes	\$0.09	8506
Copy Text from Business Card	Oscar Smith	Jan 23, 2012 (1 hour 59 minutes)	10 minutes	\$0.02	6173
Categorize Online Payments	Dolores Labs	Jan 25, 2012 (2 days 7 hours)	60 minutes	\$0.06	6083

Fig. 10. List of currently available HITs.