

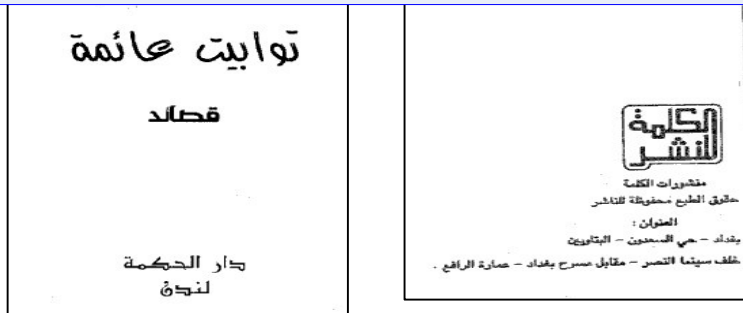
# TOWARDS CROWDSOURCING TRANSLATION TASKS IN LIBRARY CATALOGUING

A pilot study by the University of Strathclyde, Glasgow and the National Library of Scotland

ALTHOUGH AUTOMATED TRANSLATION SYSTEMS ARE WIDELY AVAILABLE ON THE INTERNET THEY ARE FAR FROM PERFECT, AND EVEN CASUAL USE DEMONSTRATES THAT THEY CAN'T PRODUCE ROBUST, ERROR FREE TRANSCRIPTIONS OF ORDINARY TEXT. HOWEVER, IT IS BECOMING INCREASINGLY APPARENT THAT CROWDSOURCING OF TRANSLATION TASKS IS NOT ONLY VIABLE BUT, IN MANY CASES, PROVIDES RESULTS EQUAL TO MORE EXPENSIVE AND SLOWER ALTERNATIVES. THIS FEASIBILITY STUDY BEGAN AN INVESTIGATION OF THE ECONOMICS OF TRANSLATING NON-ROMAN SCRIPTS. AFTER REFINING THE MANNER IN WHICH THE TASK WAS PRESENTED, THE RESULTS SUGGEST THAT THE APPROACH COULD PROVIDE BOTH SPEED AND ACCURACY FOR THE OUTSOURCING OF HIGHLY SPECIALIST TRANSLATION TASKS.

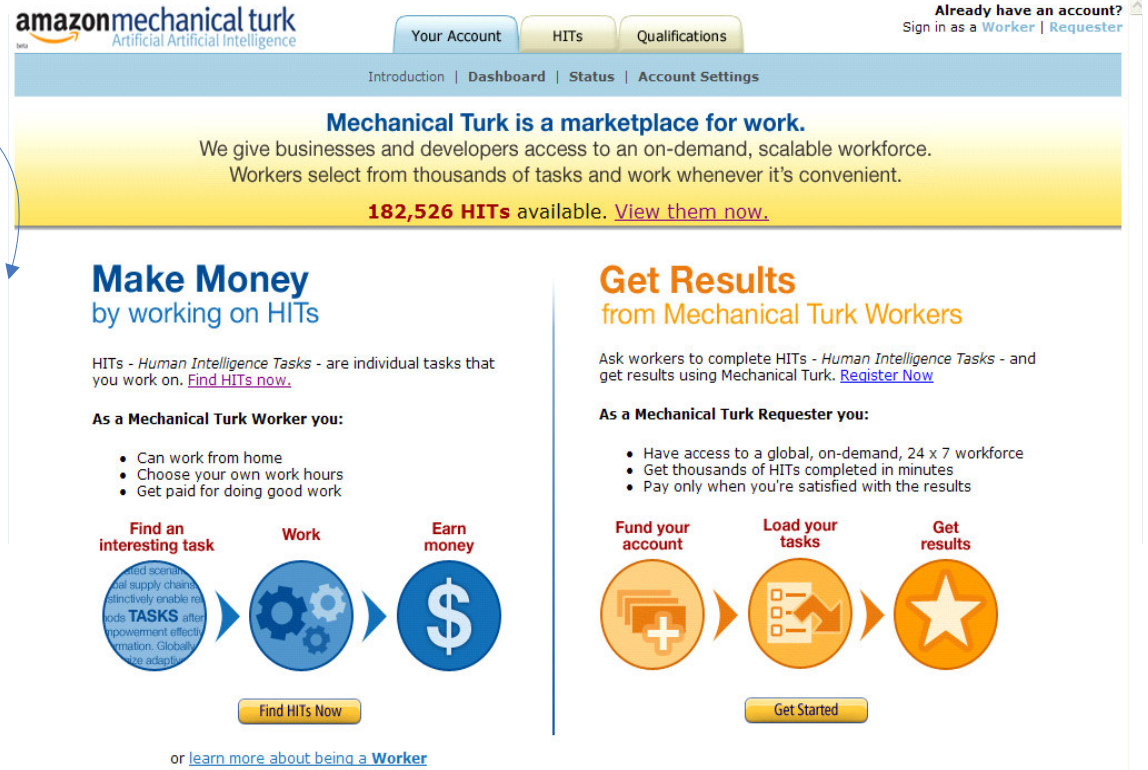
## REQUESTORS

Requestors pose human intelligence tasks (HITs): scanned images from books bearing non-roman script.



## PROVIDERS

Workers browse among existing tasks and complete them for a monetary payment set by the requester



## Amazon mTurk

A Crowdsourcing (cloud labour, micro-sourcing) marketplace that enables computer programs to co-ordinate the use of HI to perform any tasks which can be presented on a web page.

## TASK 1

Languages not specified. Response rate lower than expected, despite the generous reward of \$1.00 being offered.

## ITERATION

Recommended for more robust results, eg 1<sup>st</sup> iteration: Translation of Mandarin; 2<sup>nd</sup> iteration: Choose the best translation of this Mandarin text from among five candidates, or, edit the best translation...

## TASK 2

Languages specified. Response rate improved.

## MODERATION

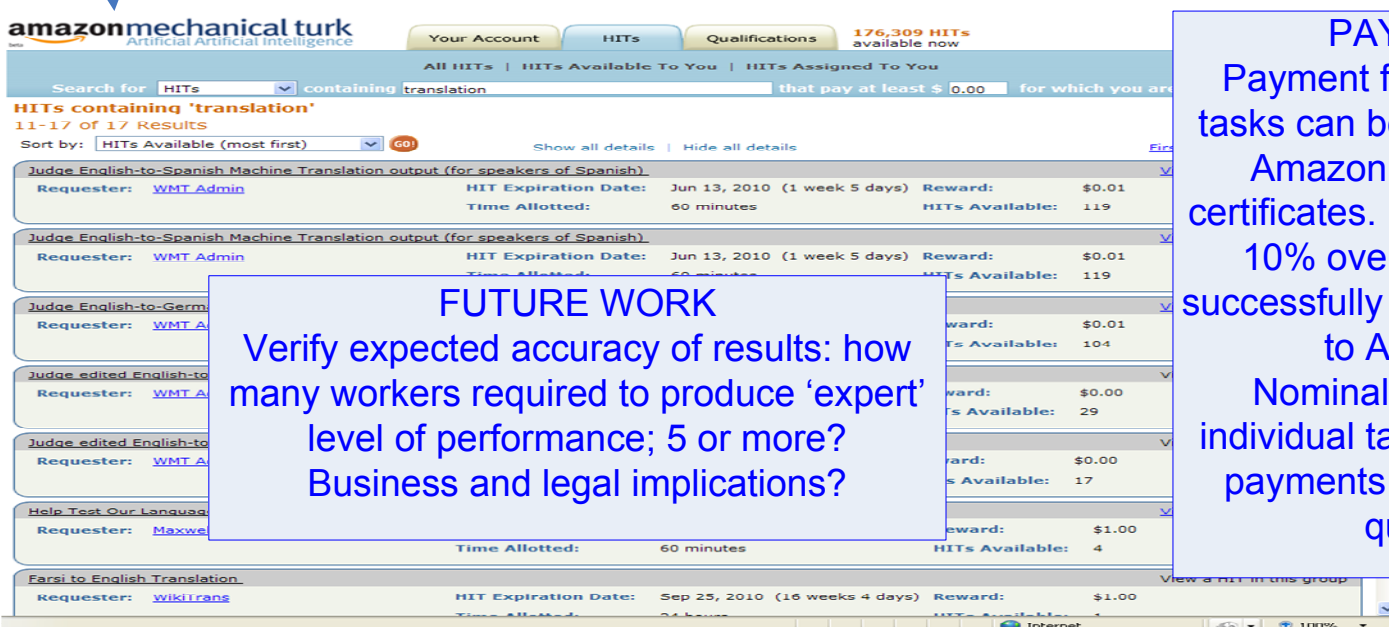
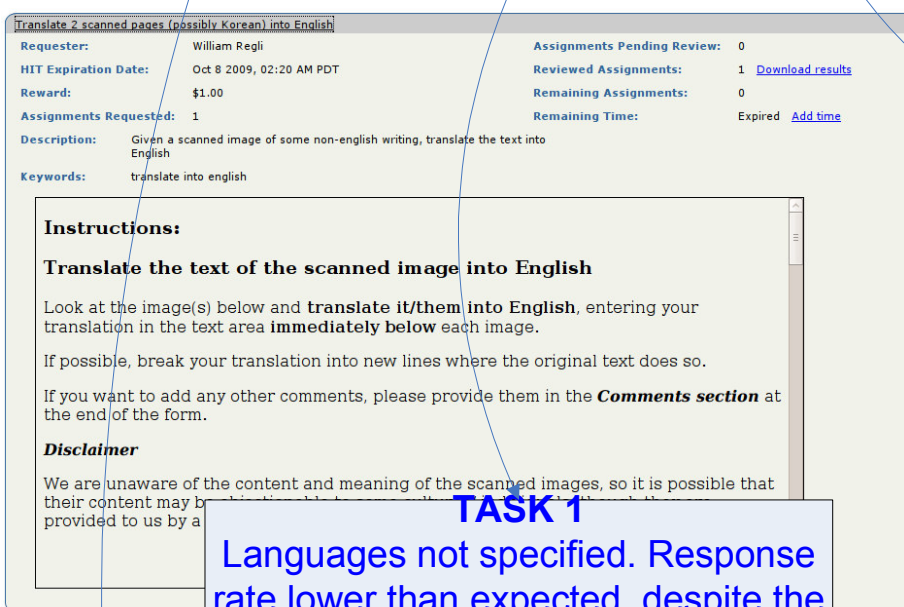
Some individuals attempt to 'game' system by supplying inaccurate data in the hope of receiving payment. However, this is likely to diminish as cohort of dependable translation workers establish mTurk reputations. Selection and merging of responses: minimum of three responses to obtain usable results & allow for flexibility of language interpretation & transliteration.

## PAYMENT

Payment for completing tasks can be redeemed on Amazon.com via gift certificates. Requestors pay 10% over the price of successfully completed HITs to Amazon. Nominal payments for individual tasks. Additional payments / bonuses for quality

## FUTURE WORK

Verify expected accuracy of results: how many workers required to produce 'expert' level of performance; 5 or more? Business and legal implications?



## CONTACTS

Dr Jonathan Corney, Department of Design, Manufacture & Engineering Management, University of Strathclyde, Glasgow.

Email: jonathan.corney@strath.ac.uk

Graeme Forbes, Head of Acquisition & Description, National Library of Scotland, Edinburgh.

Email: g.forbes@nls.uk