

Crowdsourcing African Heritage



Hussein Suleman
hussein@cs.uct.ac.za

University of Cape Town
Department of Computer Science
Digital Libraries Laboratory @ Centre for ICT4D

April 2016



Quick Outline

- Crowdsourcing
- 4 Projects
 - Malaria.net
 - Save My Heritage
 - Transcribe Bushman
 - Translate isiXhosa
- Some Thoughts

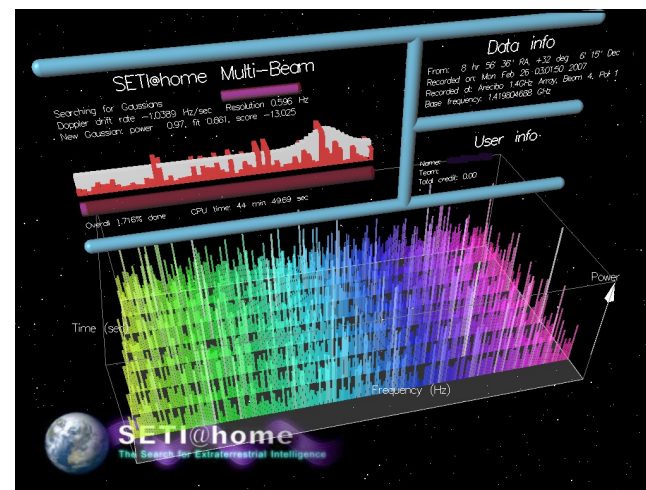


Crowdsourcing

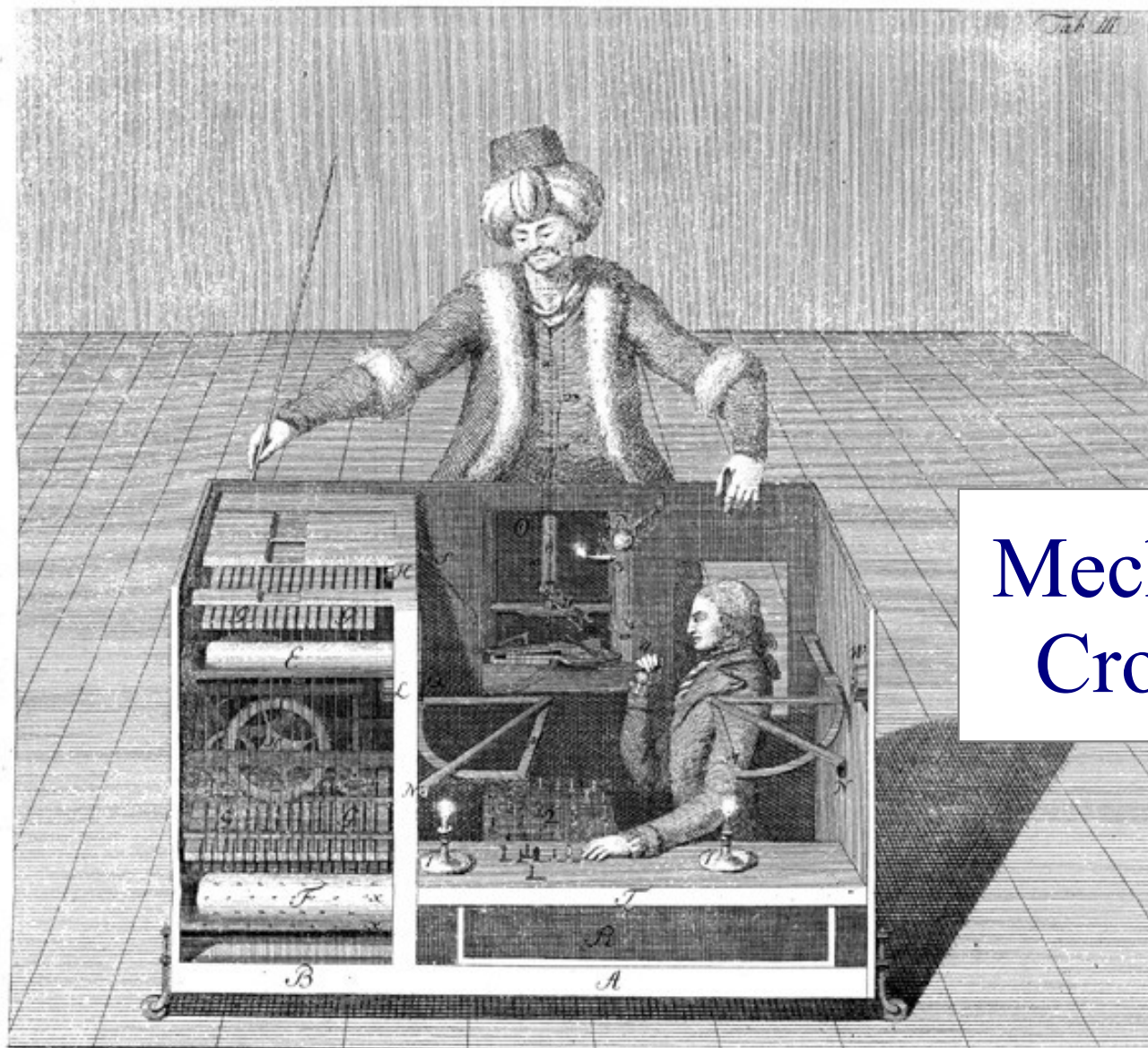
- Solving problems using a crowd of people
 - aka Human computation tasks
 - aka Volunteer thinking

- Platforms

- Mechanical Turk
 - Bossa



- Related to volunteer computing ...



Mechanical Turk – Crowdsourcing?



Is this photo showing a melanoma?

Task: 774



Classification

1. Asymmetry: ☹
2. Borders: ☹

Color: Is color uneven?

✓ Yes

✗ No

I don't know

PyBossa Demo

Home

Community

Do you see a human in this photo?

Task: 62



Yes

No

I don't know

Volunteer Thinking





Why Crowdsourcing?

- ❑ Cheaper!
- ❑ Scalable
- ❑ Can solve difficult problems
- ❑ Citizen participation





MalariaControl.net

malariaccontrol.net

Join malariaccontrol.net

1. [Read our rules and policies](#)
2. [Download BOINC](#)
3. When prompted, enter <http://www.malariaccontrol.net/>

Returning participants

- [Your account](#) - view stats, modify preferences
- [Teams](#) - create or join a team
- [Applications](#)
- [Server status](#)

Community

- [Participant profiles](#)
- [Message boards](#)
- [About us](#)

Project totals and leader boards

- [Top participants](#)
- [Top computers](#)
- [Top teams](#)
- [Other statistics](#)



User of the day



Biggles

I'm from Glasgow, Scotland. I've been into distributed computing for several years. It became an addiction when I joined...

News

Why there are no work units at the moment

Dear malariaccontrol.net community,

Apologies for not responding sooner to the lack of work-units for malariaccontrol.net. We had some issues with our attempts to fit improved models of malaria dynamics to field data. This meant that we couldn't generate new work units.

We hoped that this would be a very short term problem, but over Christmas and the New Year we experienced exactly what was going on. Unfortunately we are now faced with the challenge that for a variety of reasons, no team members able to address these problems are currently not available, and we don't have the resources that we can allocate to hiring anyone new.

We are now almost at the 10th anniversary of malariaccontrol.net and we are really hoping to get the system up and running probably be another month or two before we can get the system up again.

Malariaccontrol.net has now been going for ten years, and your contributions to this over the years have been very useful- (take a look at <https://github.com/SwissTPH/openmalaria/wiki>). Instead of this apology, we would really like to be sending you all a big thank you! 28 Jan 2016

A thank you and some science updates

Dear malariaccontrol.net supporters,

It has been some time since we sent you a science post, not because we haven't been busy, but because we've been so busy with the system.

But before sharing exciting science, we'd like to express our deep gratitude for your continued support and computational time.

Thank you!





MalariaControl.net

- ❑ Modelling of transmission of malaria
- ❑ Volunteer computing model
- ❑ Using BOINC
 - In collaboration with Swiss Tropical Institute, CERN, Berkeley
 - Famous for SETI@Home
- ❑ Server in South Africa in 2008 to support African problems





MalariaControl.net

▣ Lessons Learnt:

- Not all that useful to have a local server
- Main project was based elsewhere anyway
 - ▣ All we could run was tests
- Still many participants who volunteered cycles for tests!
 - ▣ Though not necessarily local participants





Save My Heritage

Save My Heritage

This application forms part of research on the preservation of heritage. [More Information...](#)

Full Collection

- Full Collection
- My Collection
- Add a Picture

leaderboard

#1	E	_____
#2	J	_____
#3	M	_____
#4	F	_____
#5	M	_____
#6	M	_____
#7	J	_____
#8	J	_____
	...	

Search

aerial photography africa
angola animal beach
boat botswana
braai building
cape of good hope
cape point
cape town chewa
christmas
clifton desert
drakensberg dubai
flowers
france history holiday
ireland kirstenbosch
lighthouse
lions head luanda markets



Apple Cape Town Houtbay

Add a Tag

page 1 of 36





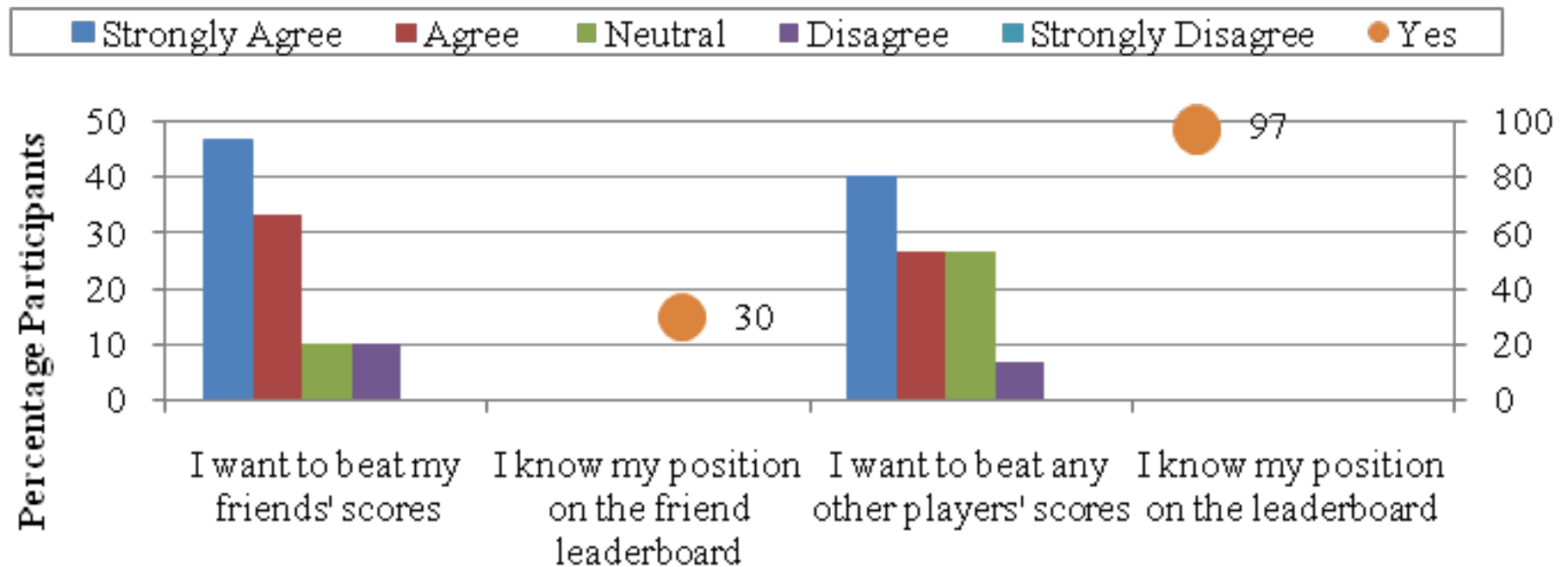
Save My Heritage

- ▣ Crowdsourcing to build a heritage image collection.

- ▣ Facebook App
 - Submit images and view/search images
 - Tag cloud for navigation
 - Using notifications to get more users
 - Both badges and leaderboard



Save My Heritage





Save My Heritage

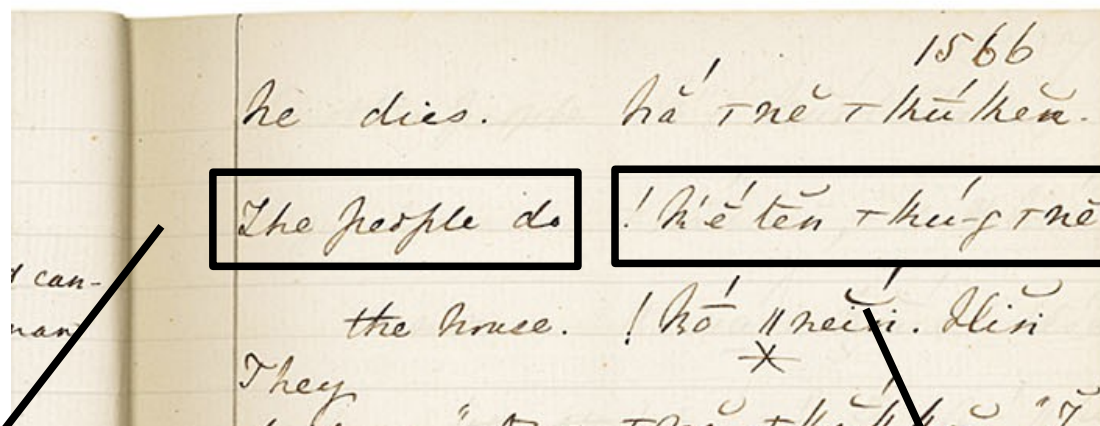
□ Outcomes:

- Competition is the norm rather than collaboration.
 - Users preferred leaderboards to badges.
 - Global domination is preferable!

- Users are not as keen to contribute as expected.
 - It does not seem to matter that there is gamification, or that the platform is a popular one.



Transcribe "Bushman"



The people do

!h'e'ten T|hú-p Tne

Williams ... AI for transcription, 45% accuracy!

Bossa/PyBossa

- ❑ Framework for volunteer thinking projects.
- ❑ Built on BOINC for job management.
- ❑ PHP/Python hooks for tasks to be performed.
- ❑ Mostly Web-based.
- ❑ Has mechanisms to create new jobs, manage jobs, etc.







Transcription Application 2/3

xóä'xóä - To Write

```
\textdoublepipe{}ne\uuline{i}n  
a\u{u}h\uuline{i}n  
t\uuline{a}tt
```

Enter English Text Here...

Convert to Latex

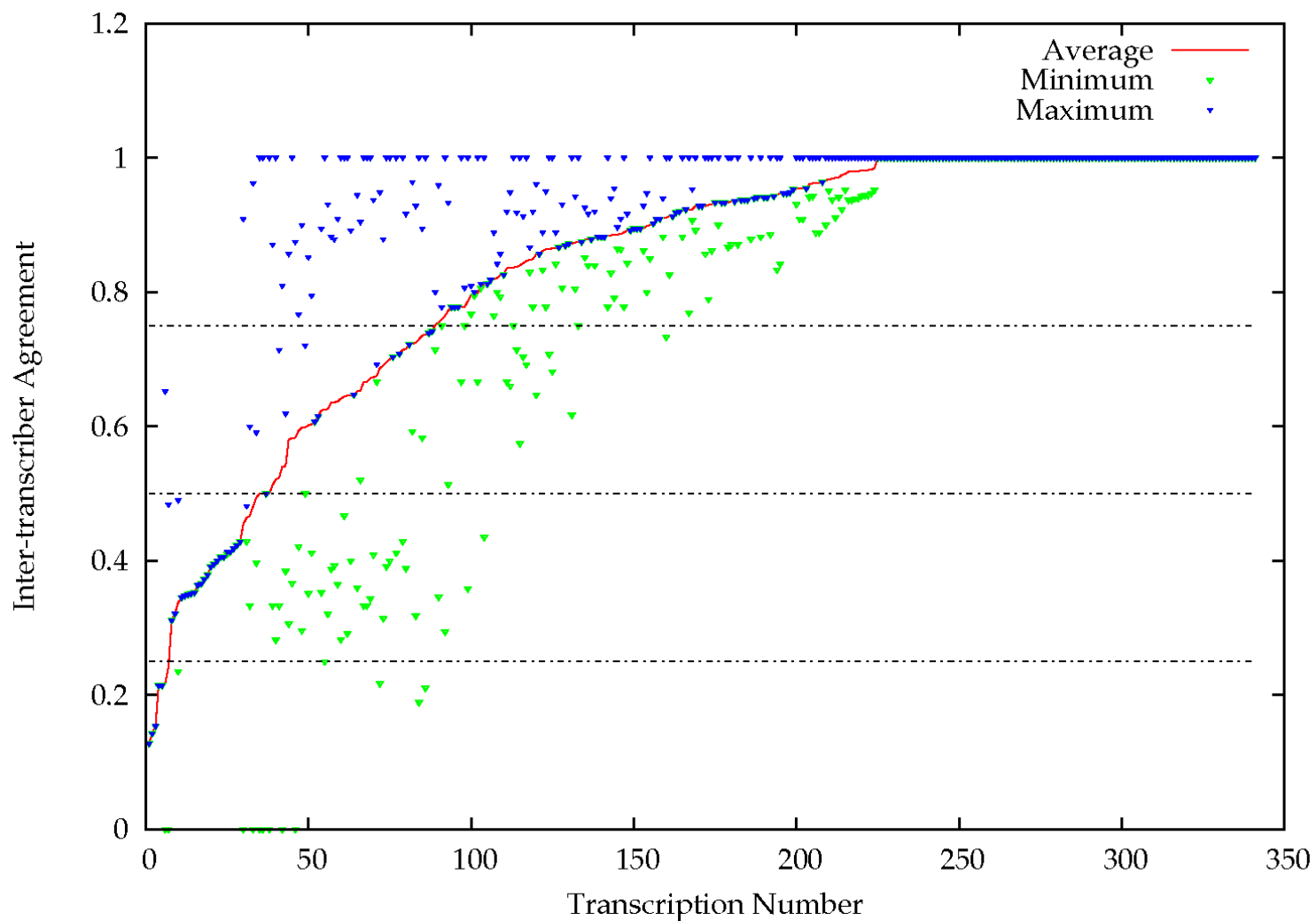
Characters and Diacritics

Diacritics Above

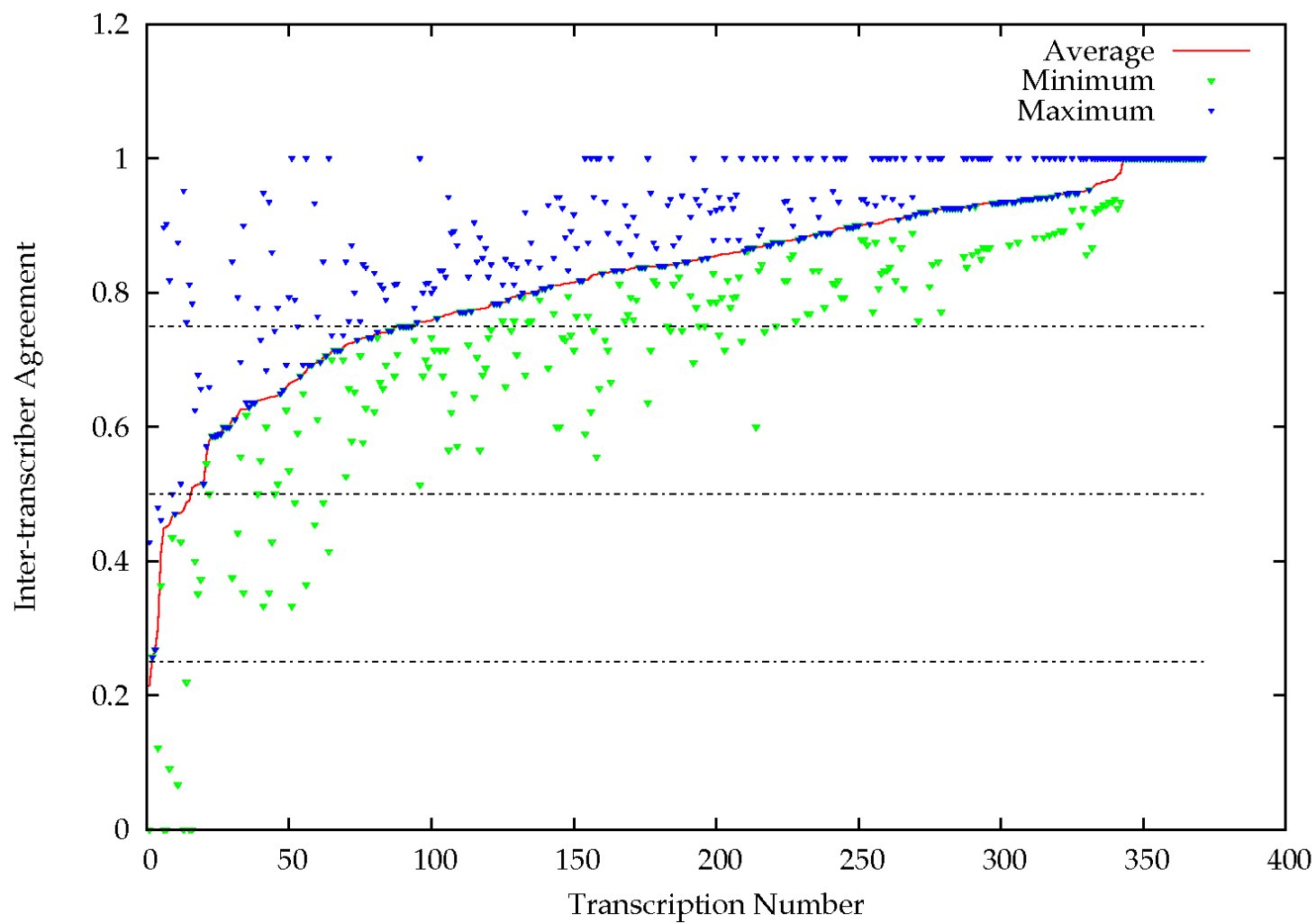




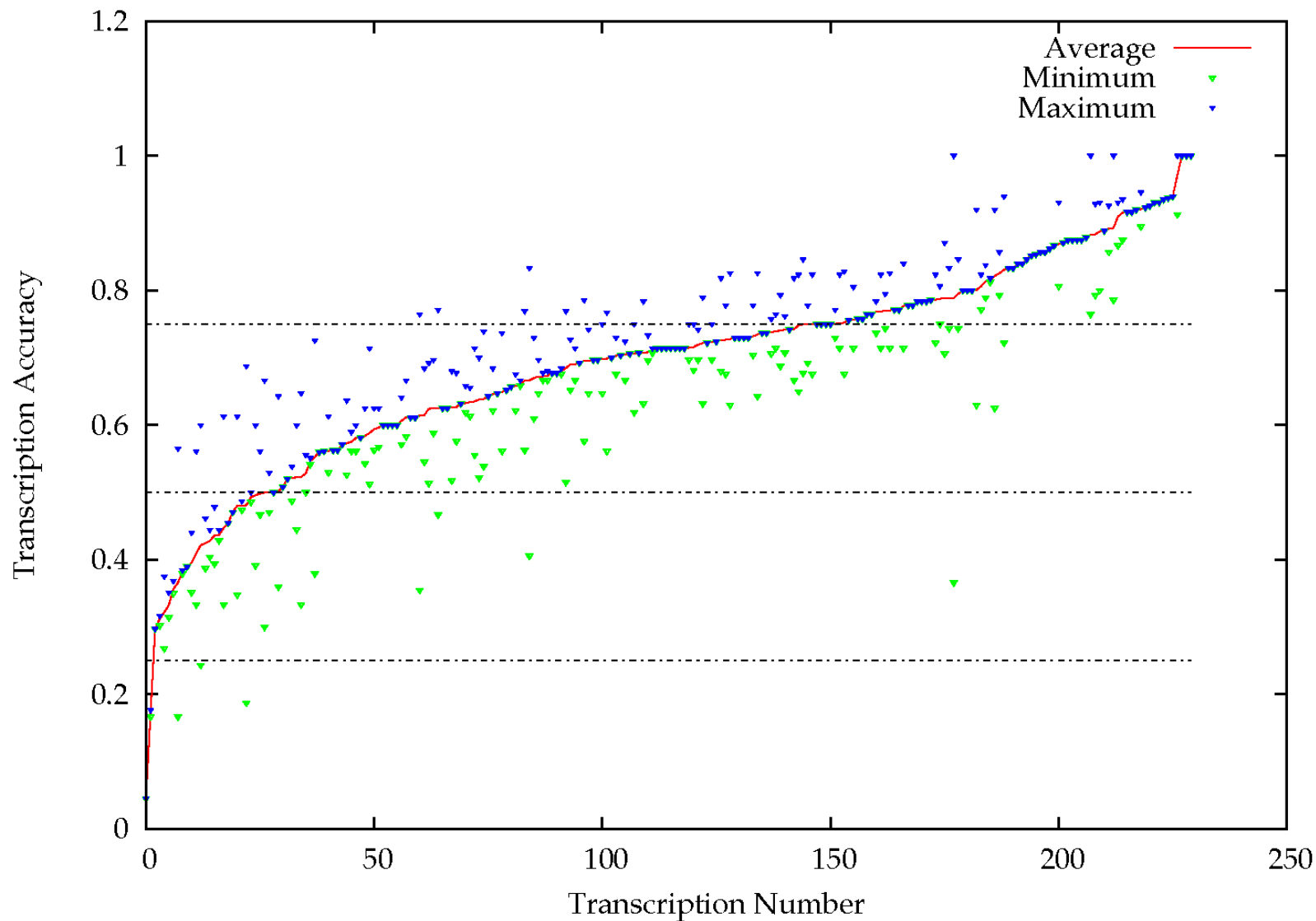
Inter-transcriber Agreement English



Inter-transcriber Agreement | Xam

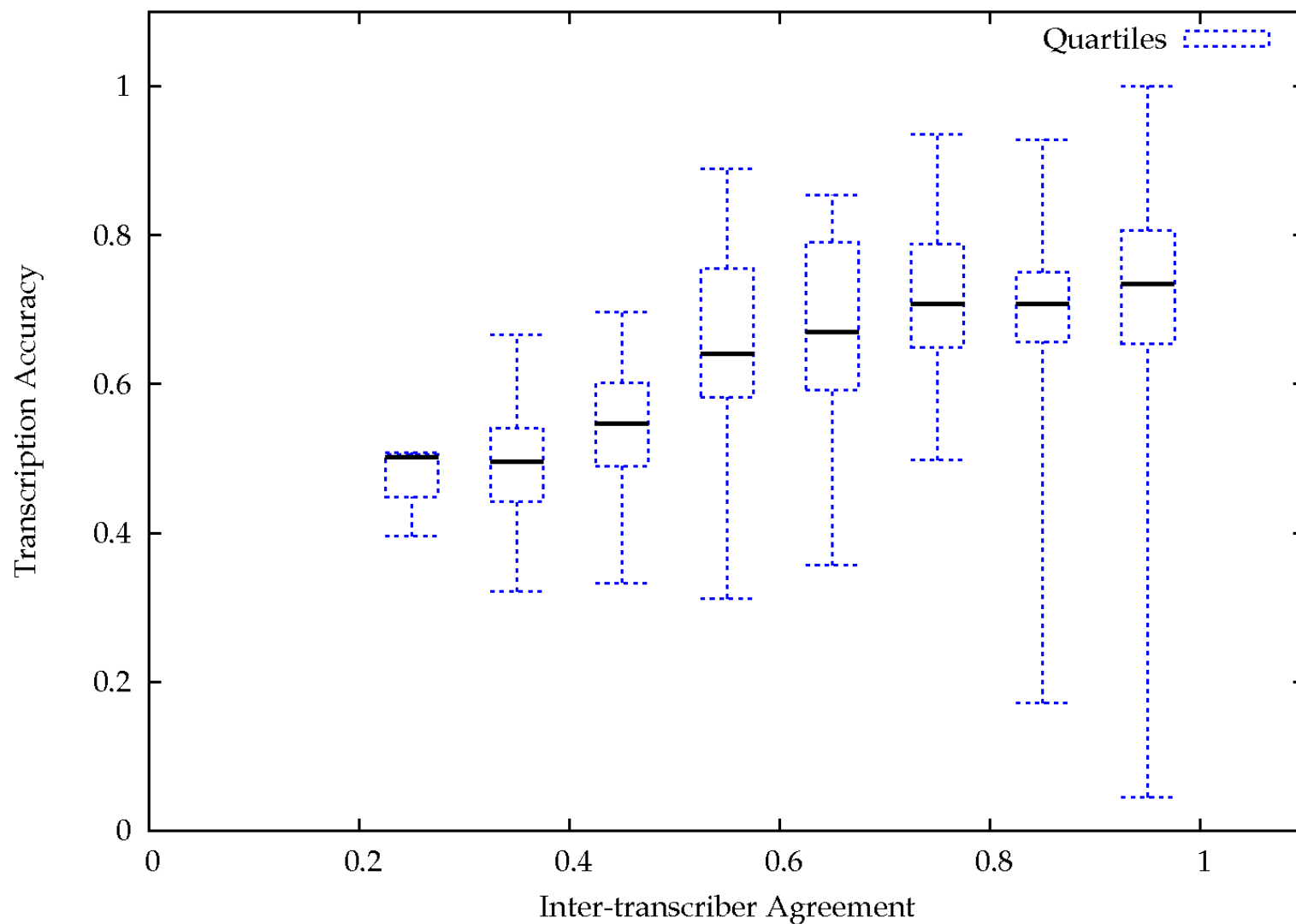


Transcription Accuracy





Correlation





Transcribe "Bushman"

□ Outcomes:

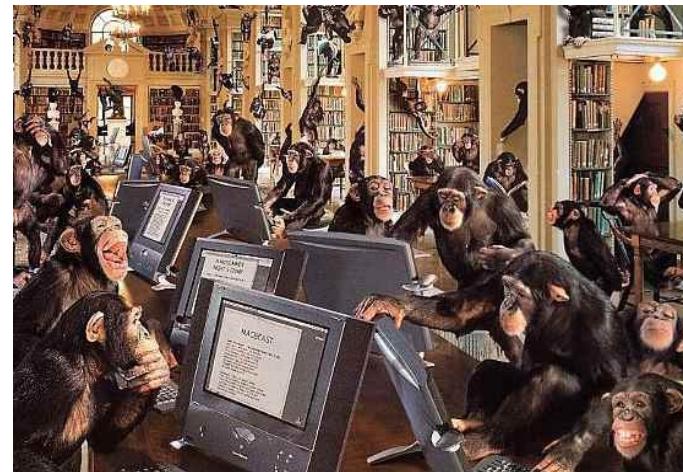
- Correlation between how much users agree and quality of transcriptions!
- Agreement as a proxy for accuracy.
- Can improve accuracy dynamically by getting more data to improve agreement.



Transcribe "Bushman"

□ Open Issues:

- Volunteer recruitment
- Answer merging of multiple answers
- Volunteer quality?





isiXhosa Translate

- ❑ isiXhosa is a low resourced language with limited digitised texts and language corpora, which are needed for building information retrieval services such as search and translation.
- ❑ *Goal: Gather original and relevant multilingual content for assembling language corpora by crowdsourcing translations via an online competitive game.*





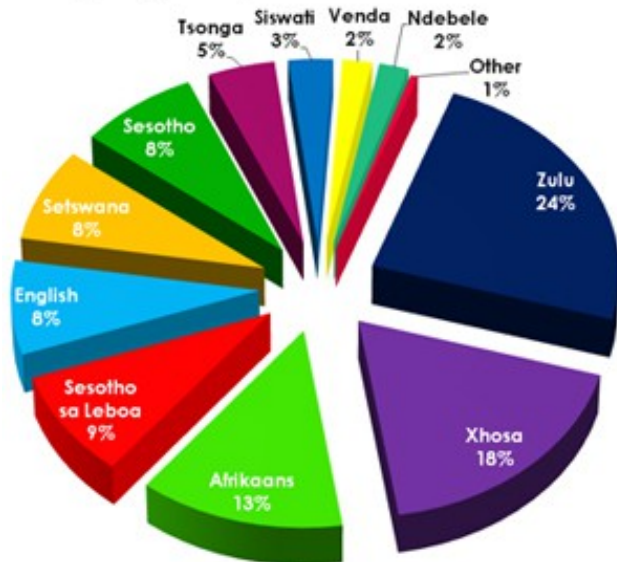
The isiXhosa Language

English	isiXhosa (Xhosa)
Welcome	Siya namkela nonke
Hello (General greeting)	Molo (sg) Molweni (pl)
How are you?	Unjani? (sg) Ninjani? (pl)
Reply to 'How are you?'	Ndiyaphila, enkosi. Kunjani kuwe? Ndiphilile enkosi, unjani wena?
Long time no see	Molo (sg) Molweni (pl)
What's your name?	Ngubani igama lakho? Igama lakho?
My name is ...	Igama lam ngu ...
Where are you from?	Uvela phi na?
I'm from ...	Ndivela e...
Pleased to meet you	Ndiyavuya ukukwazi
Good day	Ube nemini emnandi
Good afternoon (Afternoon greeting)	Uben' emva kwemini entle
Good evening	Ukhusuka shule



How many isiXhosa speakers?

Languages spoken in South Africa





Experiments

- Used a custom crowdsourcing system
- Investigated if intrinsic motivation or gamified motivation could influence users, with monetary payments being only secondary.
 - 2 experiments appealed to the intrinsic value of the task.
 - 2 experiments offered payments, but were gamified to test whether the game elements were more appealing than financial reward.





Experiment 1: System

0 volunteers have contributed 0 translations

Help us reach 100 volunteers by translating now!

This project is part of research being conducted at the University of Cape Town's (UCT) Computer Science Department by Sean Packham and Professor Hussein Suleman. Volunteers are needed to translate English Wikipedia articles into isiXhosa. Your efforts will directly make more isiXhosa content available online and help us build online services for isiXhosa speakers.



PLEASE TRANSLATE THE ENGLISH TEXT BELOW INTO ISIXHOSA. PLEASE PROVIDE YOUR EMAIL ADDRESS IF YOU WOULD LIKE TO BE CONTACTED TO PARTICIPATE FURTHER.		EMAIL <input type="text"/>
ENGLISH TEXT Is there someone here who speaks English	ISIXHOSA TRANSLATION <input type="text"/>	
ENGLISH TEXT The earliest known remnants in the region were found at Peers cave in Fish Hoek and date to between 15,000 and 12,000 years ago.	ISIXHOSA TRANSLATION <input type="text"/>	
		SUBMIT





Experiment 1: Methodology

- ❑ A pilot project to find out if participants could be gathered from Twitter.
- ❑ People were asked to contribute voluntarily.
- ❑ Curated tweets that highlighted the intrinsic value of the project were used.
- ❑ Relevant hashtags were used, such as: #isiXhosa, #Xhosa, #UCT, #SouthAfrica, #crowdsourcing





Experiment 2,3,4: Translate

[Home](#)[Leaderboard](#)[Translate](#)[Rank](#)[Profile](#)[Log Out](#)

Score: 2244

R 20.00

English Text

[110] This was confirmed after his death by the SACP and the ANC. According to the SACP, he was not only a member of the party, but also served on the party's Central Committee, when he was arrested in 1962 and this was denied for political reasons

Translate the English text into isiXhosa

Submit

You can also increase your score and earn money by ranking other user's translations on the [rank page](#).





Experiment 2,3,4: Rank

[Home](#) [Leaderboard](#) [Translate](#) [Rank](#) [Profile](#) [Log Out](#)

Score: 2244

R 20.00

English Text

[90] With the involvement of the South African Indian Congress, the Coloured People's Congress, the South African Congress of Trade Unions and the Congress of Democrats, the ANC planned a Congress of the People, calling on all South Africans to send in proposals for a post-apartheid era.

Rank the isiXhosa translations of the English text

A rank of 1 is the best and a rank of 3 the worst.

[90]ekuthatheni inxaxheba yoMzantsi Africa wo khongolose we India, ukhongolose wabantu bebala, ukhongolose woMzantsi Africa we trade unions kunye no khongolose we nkululeko, iANC yazama ukhongolose wabantu, ebiza wonke umntu wase Mzantsi Africa ukuba athumele

☐ 1 ☐ 2 ☐ 3

INtsona koloni yingingqi ubatuleke kakhulu ngezokhenketho, ingenisa ipesenti ezilithoba nemivo esibhozo kwingeniso yezemali zesizwe. Ivulela amathuba emisebenzi kubasebazi abagenge pecenti izilithoba nemivo esibhoza engqesho.

☐ 1 ☐ 2 ☐ 3

Indawo yokuhlala esemantla yeyona ifumene abantu abaninzi abasuka kumbindi wesixeko kwaye ezinye indawo zokuhlala zisetyenziselwa amashishini ingakumbi indawo yase Sandton , ukusuka emantla ukuya e Midrand, umda osepakathini phakathi kwe Rhawuti nesixeko sesizwe i Pitoli

☐ 1 ☐ 2 ☐ 3

Submit

You can also increase your score and earn money by translating content on the [translation page](#).





Experiment 2,3,4: Leaderboard

Home	Leaderboard	Translate	Rank	Profile	Log Out
----------------------	-----------------------------	---------------------------	----------------------	-------------------------	-------------------------

Leaderboard

#	Display Name	Score
1	Thulani	10060
2	KyLo	9797
3	Phiwe	9445
4	Zan	9404
5	Enkosi	9076
6	Siphamandla	6705
7	Lelethu	5073
8	Soso	5010
9	Soso	3160
10	Soso	1000





Experiment 2: Methodology

- ❑ Inspired by games that offer increasing rewards from increasing effort over time.
- ❑ For comparison, schemes that offered consistent and decreasing rewards from increasing effort and constant effort were also tested.
- ❑ Users were rewarded with points for translating and ranking and the number of points awarded depended on which group they were in.
- ❑ Each group had a leaderboard.





Experiment 3: Methodology

- ❑ Tested whether the same students from the University of Cape Town would contribute without any financial reward.
- ❑ Users were awarded 1 point for translating or ranking and a single leaderboard was used.
- ❑ Translation and ranking caps were removed, as there was no budget that could be exhausted.





Experiment 4: Methodology

- ❑ Pay users based on their leaderboard placement rather than contributions.
- ❑ An increasing reward for increased effort approach was adopted when choosing the payment points
- ❑ Only the top 40 positions were allocated a reward.





Experiment 1: Results

- ❑ Experiment 1 was run over 3 days.
- ❑ 5 tweets were sent to 132 followers and retweeted 4 times.
- ❑ Website was visited 10 times but no one contributed any translations.





Experiment 2: Results

- ❑ 200 students signed up, 121 made at least one contribution and 61 users contributed enough to receive a reward.
- ❑ 3600 individual translations and 2589 individual rankings were contributed.
- ❑ 1088 sentences received 3 translations and 734 sentences received 3 rankings.
- ❑ Increasing rewards, increasing effort, earlier payment resulted in more work.





Experiment 3: Results

- ❑ 47 users registered, 12 made at least one contribution.
- ❑ The activity of the users was considerably lower than that of Experiment 2: the most active user contributed 11 translations and 2 rankings.
- ❑ Only 11 sentences were translated 3 times and 2 sentences were ranked 3 times.
- ❑ Offering a monetary reward was considerably more successful at attracting and engaging participants.





Experiment 4: Results

- ❑ 147 users: 57 users contributed at least one translation or ranking.
- ❑ 1865 individual translations and 1767 rankings.
- ❑ 617 sentences received 3 translations and 584 sentences received 3 rankings.
- ❑ Achieved a translation cost of ZAR0.22 per word - almost double the rate of experiment 2.
- ❑ The promise of payment because of leaderboard position was as good a motivation as direct payments for tasks completed.





isiXhosa Translate

□ Outcomes:

- People (possibly from lower income communities) do not volunteer without payment, no matter how noble the cause.
- Increasing payments for increasing work.
- Participants would not contribute if payment is taken away.
- People wanted a guaranteed rate but are happily gamified using a leaderboard.





Some Thoughts

- What works in ZA:
 - Leaderboards
 - Increasing payments
 - Agreement as a proxy for accuracy
- What does not work in ZA:
 - Collaboration
 - Volunteers / intrinsic motivation
- Crowdsourcing in the heritage sector is expensive, but still cheaper than other options!





What Next?

- Are there economic class or/and other indicators for why people volunteer time/effort?
- How do we motivate users to do what they would not otherwise do for the common good?
- Can we do better than leaderboards for gamification?



questions, comments, ...



Google "hussein suleman"
Facebook/slumou
Twitter@slumou
hussein@cs.uct.ac.za

enkosi
hamba kakuhle
thank you and go well