

# **Mobile Crowd Sourced Translation for Emerging Technologies**

by

**Leroy Finn, BA (Mod.)**

A Dissertation submitted to the University of Dublin, Trinity College in  
partial fulfilment of the requirements for the degree of Master of Science in  
Computer Science

2011

# Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

.....  
Leroy Finn  
August 31, 2011

# Permission to Lend and/or Copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

.....

Leroy Finn  
August 31, 2011

# Acknowledgments

I would like to thank my project supervisor Dr. David Lewis for all his help and support during this project. It couldn't have been a success without him. I would like to thank all my family especially my parents for helping me through the Masters. Finally I would like to thank all the participants who took part in the evaluation of this project.

*Leroy Finn*

August 2011

# Mobile Crowd Sourced Translation for Emerging Technologies

Leroy Finn

University of Dublin, Trinity College, 2011

Supervisor: Dr. David Lewis

Facebook and NGOs such as Translators Without Borders, have shown the power of motivating unpaid volunteers, working over the web, to complete large scale document translation tasks. The volunteers that take part gain a sense of achievement and community with the other volunteer translators. It also gives these volunteers that take part a valuable training experience for those that are trying to develop their professional translation skills. At the moment, most crowdsourced translation systems on the market are designed and targeted at desktop based browsers inspired by the design of professional computer assisted translation tools. This means a large population of potential translators cannot access these tools because these translators are living in less developed or emerging economies where PC access is rare. In these less developed economies, Internet access for business and personal computing is mostly conducted via mobile phones. In some countries, even where economic development is strong a majority of people don't choose desktop computing for conducting their personal computing business but opt for their mobile phones.

This dissertation examines by means of a prototype, how translation crowdsourcing could be offered in such regions by providing powerful services (such as document segmentation, community management and machine translation) on the cloud in a way that can be accessed via web browsers on mobile devices. This prototype is unique in the way that translation data is stored semantically via the use of a RDF triple store. This means relationships can be built up between data in the triple store such that data can be captured, shared and reused in many other jobs. This means that the language resources in the triple store could help to ease issues that exist in long-tail localization jobs, where translation of material is needed but the demand is not big enough to warrant a translation. The prototype was evaluated on a sample crowd to see if this type of application is usable on a mobile device and also to see if the prototype can truly harness the power of the crowd.

# Contents

Mobile Crowd Sourced Translation for Emerging Technologies .....	1
Declaration.....	2
Permission to Lend and/or Copy.....	3
Acknowledgments.....	4
Abstract.....	5
Contents .....	6
Table of Figures .....	10
<b>Chapter 1 Introduction.....</b>	<b>12</b>
1.1 Background.....	12
1.2 Motivation.....	12
1.3 Hypothesis.....	13
1.4 Research Approach .....	13
1.5 Outline of Dissertation.....	14
<b>Chapter 2 State of The Art.....</b>	<b>15</b>
2.1 Translation and localisation .....	15
2.2 What is Mobile translation?.....	16
2.3 Crowdsourcing.....	17
2.4 Crowdsourcing Techniques.....	21
2.5 Open issues in mobile translation .....	27
2.6 Technology Research.....	28
2.7 Conclusion .....	35
<b>Chapter 3 Design.....</b>	<b>38</b>
3.1 Functional Requirements .....	38
3.1.1 Sociability .....	38
3.1.2 Usability.....	38

3.2 Design Choices.....	40
3.2.1 Project Scoping.....	39
3.2.2 Data Storage.....	39
3.3 Key Components.....	41
3.3.1 User Pages.....	41
3.3.2 Job Life Cycle.....	41
3.3.3 Deliberations.....	43
3.3.4 Translation.....	44
3.3.5 Keyboards.....	45
3.3.6 Data Model.....	45
3.3.7 RDFLogger.....	48
3.3.8 Mobile design.....	49
<b>Chapter 4 Implementation.....</b>	<b>52</b>
4.1 Deployment.....	52
4.1.1 Sesame Server.....	53
4.1.2 Sesame Workbench.....	53
4.1.3 RDFLogger.....	53
4.1.4 Visualiser.....	54
4.1.5 Prototype.....	56
4.2 Details of Implementation.....	56
4.2.1 Home Page.....	56
4.2.2 Login Page.....	57
4.2.3 Registration Page.....	58
4.2.4 User page.....	59
4.2.5 Job Page.....	61
4.2.6 Editor Page.....	65
4.2.7 Publisher Page.....	70
4.2.8 Reputation board.....	73

<b>Chapter 5 Evaluation</b> .....	74
5.1 Evaluation Approach .....	74
5.1.1 Evaluation Requirements .....	74
5.1.2 Evaluation Scoping .....	74
5.1.3 Fake data .....	75
5.1.4 Evaluation process .....	75
5.1.4.1. System Usability Scale (SUS) .....	76
5.1.4.2. Collaboration Analysis.....	77
5.1.4.3. Application Analysis.....	77
5.2.4.4. Participant Profile .....	78
5.2 Pilot Study.....	78
5.2.1 Pilot Study Findings.....	78
5.2.1.1. Pilot Study System Usability Scale.....	79
5.2.1.2. Pilot Study Collaboration Analysis.....	80
5.2.1.3. Pilot Study Application Analysis.....	81
5.2.1.4. Pilot Study Participant Profile .....	82
5.3 Pilot Study Feedback .....	83
5.3.1 Amendments .....	84
5.4 Main Study.....	86
5.4.1 Main Study Findings.....	86
5.4.1.1 System Usability Scale .....	86
5.4.1.2 Collaboration Analysis.....	90
5.4.1.3 Application Analysis.....	94
5.4.1.4 Participant Profile .....	97
5.4.2 Feedback .....	99
5.5 Evaluation summary .....	100
<b>Chapter 6 Conclusion</b> .....	102
6.1 Conclusion .....	102
6.2 Future Work.....	103
Appendix A MySQL.....	105



Appendix B Questionnaire..... 108

Appendix C Data Collected ..... 112

Appendix D Worksheet..... 116

Bibliography ..... 117

# Table of Figures

Figure 2.1: Purposes why individual use community translation.....	18
Figure 2.2: Crowdsourcing challenging issues .....	19
Figure 2.3: An example of Facebook crowdsource translation.....	20
Figure 2.4 : Crowdsourcing vs Machine Translation .....	21
Figure 2.5: Ankits Model .....	26
Figure 2.6: Other mobile crowdsourcing projects .....	27
Figure 2.7: XLIFF example .....	30
Figure 2.8: An example RDF graph .....	32
Figure 2.9: Popular Mobile Web browsers .....	35
Figure 3.1: Sesame Vs. Jena .....	40
Figure 3.2: Job Life Cycle .....	41
Figure 3.3: Job Management.....	41
Figure 3.4: Job Contribution .....	42
Figure 3.5: Sample RDF graph data from Prototype.....	46
Figure 3.6: MySQL data model.....	47
Figure 3.7: Examples of badly formatted websites.....	49
Figure 3.8: Examples of correctly formatted websites .....	50
Figure 3.9: Viewport META TAG .....	50
Figure 4.1:Deployment .....	52
Figure 4.2: Sesame server.....	53
Figure 4.3: Visualiser.....	55
Figure 4.4: Prototype Architecture.....	55
Figure 4.5: Home page.....	57
Figure 4.6: Login page.....	58
Figure 4.7: Registration page .....	59
Figure 4.8: User page image .....	61
Figure 4.9: Sentence type in XLIFF.....	64
Figure 4.10: Job creation page .....	64
Figure 4.11: URI sample .....	65
Figure 4.12: User Contribution being added to the XLIFF document .....	69

Figure 4.13: Editor page.....	70
Figure 4.14: Publisher page .....	71
Figure 4.15: Publishing Job XLIFF .....	72
Figure 4.16: Reputation board.....	73
Figure 5.1: Pilot Studies System Usability Scale Average .....	79
Figure 5.2: Pilot Studies SUS Results .....	80
Figure 5.3: Pilot Study Collaboration Analysis.....	81
Figure 5.4: Pilot Studies Application Analysis .....	82
Figure 5.5: Pilot Study Participant Profile .....	83
Figure 5.6: SUS Results .....	87
Figure 5.7: System Usability scale results for the each test subject .....	89
Figure 5.8: Average System Usability scale result .....	89
Figure 5.9: Question 11.....	91
Figure 5.10: Question 12.....	91
Figure 5.11: Question 13.....	92
Figure 5.12: Question 14.....	93
Figure 5.13: Question 15.....	94
Figure 5.14: Question 16.....	95
Figure 5.15: Question 17 .....	96
Figure 5.16: Question 18.....	96
Figure 5.17: Question 19.....	97
Figure 5.18: Question 20.....	98
Figure 5.19: Mobile browsers used in the evaluation.....	99

# Chapter 1

## Introduction

### 1.1 Background

Localization and translation of data can be a labour-intensive and often requires significant efforts from development teams. There are tools that can simplify the localization/translation process such as SDL Trados [57], Alchemy Catalyst [1] and WeLocalise GlobalSight [62]. Typically content creation companies outsource to specialized companies, called Language Service Providers, to cut costs. An emerging alternative for translating content is crowdsourcing. One major advantage of using crowdsourcing is that the assessment of the crowd is more in-depth cause of the volume of translators and is free compared to having to pay professional translators to do the work. This is majorly advantageous if the people rating the translations are the users of the product or service as they are the ultimate end users of the product and this will result in more usable language being used in the translation/localisation.

In the last few years many bilingual users are being motivated by social aspects to take part in this localization process such as in Facebook which may be referred to as collaborative or social translation. This social translation allows for translation of content to be completed via a large body of translators, whether they are professional linguists, amateurs or volunteers

### 1.2 Motivation

Companies such as Facebook and NGOs such as Translators Without Borders, have shown the power of motivating unpaid volunteers, working over the web, to complete large scale document translation tasks. As well as the volunteers gaining a sense of achievement, and often a sense of community with other peer translators, it can also provide valuable training experience for those developing their professional translation skills. Most crowdsourced translation systems are designed for desktop based browsers inspired by the design of professional computer assisted translation tools used by professional translators [17].

This makes them inaccessible to a large population of potential translators residing in emerging economies, where access to a PC is rare [20]. Indeed in many such countries, even as economic

development proceeds, people do not opt for desktop computing as incomes rise, but find themselves able to conduct their personal computing entirely via their mobile phones [73]. Even in less developed countries, business and personal computing and Internet access is conducted largely through mobile phones. This project therefore will examine how a translation crowdsourcing application could be offered in such poor economic regions via mobile device web browsers.

### **1.3 Hypothesis**

This dissertation seeks to design an appropriate and efficient crowdsourcing translation application that can be offered globally to its users via mobile web browsers as well as desktop web browsers. This would allow the application to be accessible to a wider range of potential translators in rich and poor economic regions. The translation data in the application will be stored in an RDF triple-store which uses semantic web mechanisms which gives the application the power to represent, combine, and share knowledge between communities of distributed machines [21].

### **1.4 Research Approach**

#### **1.4.1 Scoping research**

The application created in this dissertation will focus on the context of a mobile crowd sourcing application and will not attempt to evaluate issues related to crowd-motivation due to lack of access to an existing crowd translation community or the time and resources to assemble one.

#### **1.4.2 Application features**

Crowdsourcing mechanisms will involve server side mechanisms for keeping track of who contributed what and what is yet to be completed, as well as separating out translations, review/post-editing and rating of translations. All data that is related to the translation of data will be stored in an RDF triple-stored and other data such as user data, comments and ratings scores will be stored in a MySQL database. Another element being developed is a solution to data entry issues for the application as some languages do not have keyboard support readily available.

#### **1.4.3 Evaluation process**

The application was evaluated on mobile browsers for its usability and collaboration features. The results gathered in this evaluation will help to highlight how usable the application is and

does this type of application ultimately suit a mobile device. The evaluation took place using 12 participants two in the pilot study and ten in the main study. The pilot study was held first and it helped to highlight any issues in the initial application so that they did not impede the results in the main study.

### **1.5 Outline of Dissertation**

- This dissertation starts off discussing information about the emergence of translation and its origins and reviews the current state of the art in the key areas deemed relevant to this project in chapter 2.
- In chapter 3, from the data gathered in chapter 2 a set of functional requirements are ascertained which will aid in the design choices as well as helping to identify core concepts in the proposed solution.
- In chapter 4, the prototype that has been implemented will be discussed in detail and the key features of the application will be discussed.
- Chapter 5, will discuss the evaluation of the prototype and the data that was collected from the evaluation will be analysed and conclusions will be drawn from these results
- Chapter 6, will evaluate possible future research based on the findings from the evaluation and how the work done on the prototype could be extended.

# Chapter 2

## State of The Art

This chapter's purpose is to outline and review the current state of the art research in the areas of localisation and translation technologies, crowdsourcing technologies, crowdsourcing techniques and localisation management tools.

### **2.1 Translation and localisation**

#### **2.1.1 Translation**

Translation is the communication of the meaning of a source-language text by means of an equivalent target-language text [36]. Translation has been around since the appearance of literature and it became popularised with the inception of the printing press. Translators have to deal with spill-over of source language idiom and usage into the target-language translation [28]. The demands of translation of documentation increased during the Industrial Revolution and this helped in the creation of dedicated translation schools and professional associations [65]. The labour intensive nature of translation has encouraged engineers since the 1940s to automate translation techniques (machine translation) [26]. The recent popularity in the Internet has helped to create a world-wide market for translation services which has help to facilitate the growth of language localization processes.

#### **2.1.2 Localisation**

Localization refers to the adaptation of a product, application or document content to meet the language, cultural and other requirements of a specific target market [67]. Localization isn't a simple task even though it is synonymous with translation of documentation and user interfaces for software, the issue that arise in localization are more complex such as for example "*Text and graphics containing references to objects, actions or ideas which, in a given culture, may be subject to misinterpretation or viewed as insensitive*" [67,68].

#### **2.1.3 Issues of Localisation and Translation research**

The adapting of software into native languages using translation allows access to services and software that would otherwise would be inaccessible to the population. This makes localisation of content an important part to incorporate into the design of software or web sites as

translations of the software can be a major value adding component in expanding the user base of products or services making localisation a potentially lucrative addition. There are two major challenges facing the localisation industry in adapting this digital content to culture, locale and linguistic environment which were discussed by CNGI[5] and are as follows:

- **Volume**-The amount of content that needs to be localised in many various languages is growing steadily especially since the inception of smart phones and computers in third world nations. This increase in demand massively outstrips the current capacity of current translation and localisation capacities. The outcome of this is that only a fraction of content that needs to be localised is actually localised and only localised into a small set of languages. This severely reduces business opportunities in regions that speak these languages but also this lack of localisation into these languages leads into a digital divide. This digital divide costs these regions potential useful services, products and information.
- **Access**- Access to digital content is new and evolving onto ever small devices as users can access data on the move. The traditional view of localisation of content assumed stable, reliable and corporate content and the localisation that took place on this content was assumed to be on a large scale, off- line and executed by professionals in the field of translation. Nowadays, this is less the case as much content on the web is perishable with frequent software updates taking place and rapidly increasing amount of user created content in various languages on various platform such as of blogs, forums, FAQ, knowledge articles, user wikis etc. Instant access and the ever changing digital content on the web requires a new development in fully automated localisation technologies.

## **2.2 What is Mobile translation?**

Mobile translation is typically a machine translation service provided on mobile devices such as PDAs, tablets or smart phones. The mobile devices must be able to use communication mediums such as SMS, 3G or WIFI to avail of these translation technologies. The major advantage of this type of translation is that it allows users to have portable real time automated translation for little or no cost.



## **Mobile Translation Technologies Past and Present**

Machine translation software for phones was launched in 2004, by Transclick which released a mobile translation software of text to text translation in both email and SMS messages via mobile devices. *“Transclick provides real-time translation capabilities for cellphones, messaging and collaboration systems”* [58]. Since its initial entry into the market Transclick have launched an Instant messenger component for mobile devices. The client that Transclick offers on mobile devices is available for Blackberry’s and J2ME smart phones for 16 different languages. Transclick is developed specifically for Java based phones and created using Java based technologies [58].

This notion of mobile translation has even spread into vast other markets such as a language learning tools for students. Early projects developed by Stanford Learning Lab in 2001 focused on Spanish vocabulary practicing and word and phrase translation along with testing the students. These initial mobile tools used a combination of both text and speech to help users input into the software. Having seen the success of the Spanish project a similar program was launch in Japanese university teaching English to students [16].

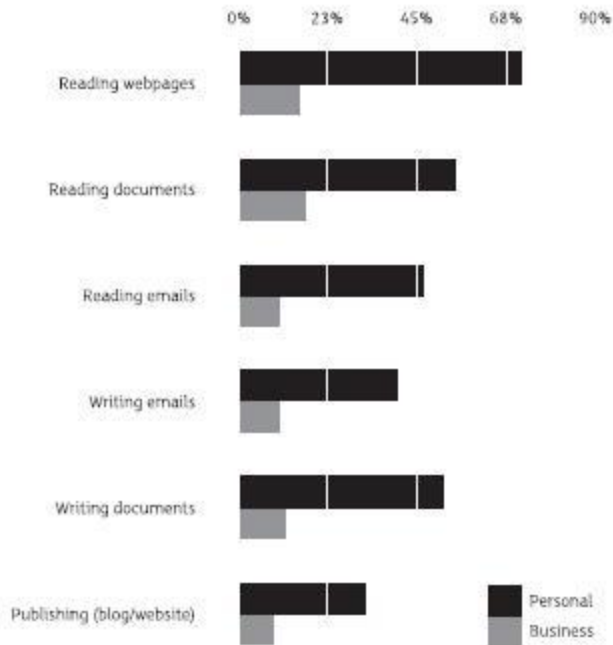
In the SMS vein companies such as Moka LCC started to offer SMS translation whereby users send a text message and get back a translation of the text taking into consideration phrases used in a particular region [56]. Since the inception of the smart phone in 2007 there has been an increase in the growth of translation software tools in the field of mobile translation such as LingoSoft Full-text Translation Software and the SYSTRAN Mobile Translator. Nowadays, though there has been a move to translation websites on mobile devices and in particularly smart phones. An application that does this is Mobile Translator which works on any mobile browser. The translation tool offers a free HTML code if users want to add a translation box onto their website [37].

## **2.3 Crowdsourcing**

Crowdsourcing has enormous potential that can be truly unleashed when extended to mobile devices, such as smart phones. The term crowdsourcing was coined in May 2006 in Wire magazine by Jeff Howe in the article “The Rise of Crowdsourcing”. Jeff stated *“Crowdsourcing is the act of taking a job traditionally performed by a designated single employee and outsourcing it to an undefined generally large group of people in the form of an open call”* [66].

The current trends in crowdsourcing revolves either around communities of participants sharing jobs openly or are published on an open marketplace where users compete to complete a task for either tender or as competition where the best translation is selected from a group of other user submissions. There have been some major success stories in the field of crowdsourcing from Mechanical Turk, GoldCorp’s challenge, and Threadless.com. Mechanical Turk uses a system of dividing up the translation project into a number of small parts and distributing these small parts across a large group of translators (Facebook and Haiti relief) [19,40].

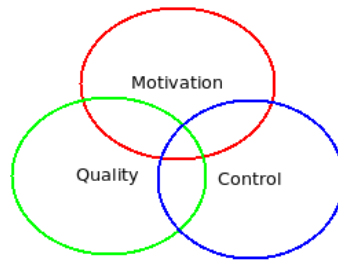
Richardson et al (2009) purposed many different pieces of content which people get translated either personally or for business purposes through crowdsourcing. The below figure 2.1 shows what people use crowdsource translation for:



**Figure 2.1: Purposes why individual use community translation [48]**

There are many advantageous aspects to crowdsourcing such as getting work done for free by a large group of participants but equally there are a range of many challenging issues such as:

- **Motivation** (money, reputation, stars etc.)
- **Quality** (work quality can be high or low)
- **Control** (security of users and content, crowd management etc)



**Figure 2.2: Crowdsourcing challenging issues [13]**

The presence of crowdsourcing has allowed for major strides in relation to assisting communities in African hospitals and Haiti etc [20,40]. Crowdsourcing has also aided in changing the business equation worldwide. It has achieved this by allowing smaller companies to be on equal footing with major conglomerates such as IBM and Microsoft because of the access to free translation skills which will potentially help smaller companies widen their user base .

### **2.3.1 Issues in crowdsourcing**

Crowdsourcing as discussed previously can be advantageous in large scale projects on the web for getting large amounts of content translated for free by the crowd. The key issue with crowdsourcing is that there actually needs to be a crowd for the job to be done efficiently and correctly. In Asia online[9] and wikibhasha [64] , there are so many eyes on the translation that the crowdsourcing model can work as users can quickly edit incorrect translations. In absence of this crowd bad translations can take place such as in the Facebook example below (Figure 2.3):

#### **English**

**Account Security** [hide](#)

To help keep your Facebook account as safe as possible, we can notify you when your account is accessed from a computer or mobile device that you haven't used before.

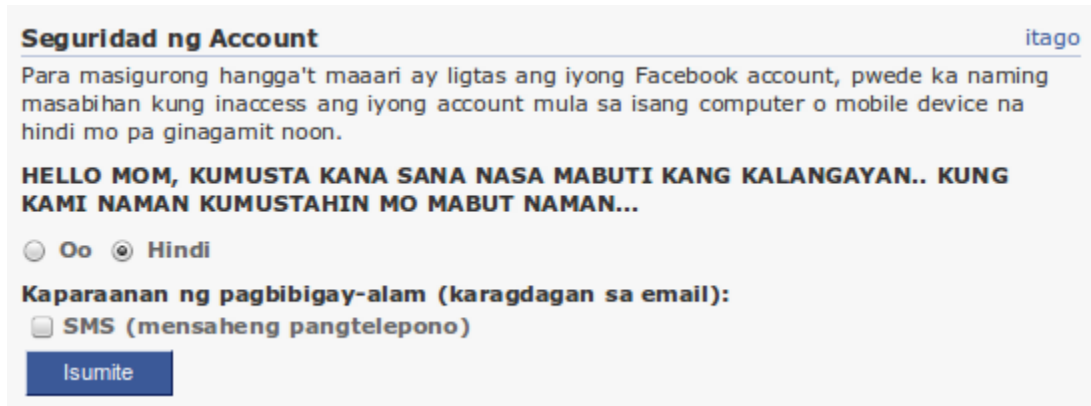
**Would you like to receive notifications for logins from new devices?**

Yes  No

**Notification methods (in addition to email):**

SMS (mobile text)

## Filipino



**Seguridad ng Account** itago

Para masigurong hangga't maaari ay ligtas ang iyong Facebook account, pwede ka naming masabihan kung inaccess ang iyong account mula sa isang computer o mobile device na hindi mo pa ginagamit noon.

**HELLO MOM, KUMUSTA KANA SANA NASA MABUTI KANG KALANGAYAN.. KUNG KAMI NAMAN KUMUSTAHIN MO MABUT NAMAN...**

Oo  Hindi

**Kaparaanan ng pagbibigay-alam (karagdagan sa email):**

SMS (mensaheng pangtelepono)

**Figure 2.3: An example of Facebook crowdsource translation [59]**

The first boldface question in Filipino translates to *“Hello Mom! How are you doing? We hope you’re doing great.”* Most native Filipino speakers prefer to use the English version of Facebook since most Pinoys have English lingual skills [59]. This example helps to highlight the golden rule of crowdsourcing to eliminate these bad translations which is to have focused managed crowds. This means the crowd needs some oversight if information is going to be translated or localized correctly into different languages. The crowd also should be working towards a goal or reward to encourage them to do correct translations. In the example above, people in the Philippines are not really encouraged to help translate Facebook into their native tongue as many people living in the Philippines can speak English well enough to use the English version [59,63].

The advantages of crowdsourcing outweigh the disadvantages such as how crowdsourcing has aided in the translation of large amounts of digital content for free in a relatively short period of time. Mobile translation crowdsourcing has particularly shown its effectiveness in blood donation banks in Africa[20] and in the wake of the earthquake crisis in Haiti [40] .

### **2.3.2 Machine translation vs. Crowdsourcing translation**

Machine Translation (MT) is sub-category of computational linguistics where a computer program analyses a source piece of text from one language to another [13]. The major issue facing machine translation is that it is far less accurate compared to human translation. In some cases though human intervention can take part in the form of post-editing of machine

translations. In the future with the advancements in machine translation, translators will become proofreaders for the machine translations.

<b>Crowdsourcing</b>	<b>Machine Translation</b>
Humans (emotionally attached) perform the translation	Computer software perform the translation
Keeps translators (professionals and non-professionals) into business	Replaces translators
No terminology consistence	Terminology consistence
Company profits	User profit
Personal content (Web 2.0)	Enterprise and valuable digital content

**Figure 2.4 : Crowdsourcing vs. Machine Translation [13]**

#### **2.4 Crowdsourcing Techniques**

Whilst mobile phones were developed for Western business executives a large majority of mobile phone subscribers live in developing countries in Africa and Asian nations. Eastern Africa for example is the fastest growing mobile phone market in the world [20]. This mainly down to the fact unlocked GSM phones in these regions of Africa with a sim card and airtime can be purchased for US\$20 or less. The expectation from these customers is not just the ability to use the mobile device as a two way communication device but also as a music player, a digital wallet and a portable web surfing device. The users in these Africa and Asian nations primarily surf the web via their mobile devices making them an essential part of everyday life and work [20].

This has of course lead to a natural situation of mobile crowd sourcing which is a new paradigm for utilizing the power of the crowd to facilitate large scale tasks that are costly or time consuming compared with traditional methods. A large amount of the world population carries their phones with them the entire day giving them the opportunity to contribute at any time to

the translation process. This can be seen to be particularly advantageous in nations that have heavy mobile web usage and could potentially speech rarer dialects such as in nations in Africa and Asia [2].

The implementation of txteagle application in remote areas of Africa has helped in highlighting a few key practices arising around crowdsourcing such as:

- Incentivisation of the crowd
- Localisation and Sentiment analysis
- Quality control
- Inferring ‘Accuracy’

### **Incentivisation of the crowd**

This is a key aspect of any crowdsource translation initiative because to attract and compel the members of the crowd to use their free time there has to be some sort of incentive for the user. These incentives can take many forms be it payment or a function that can aid them in their everyday work or maybe something trivial like getting their friends to join on that website. So there has to be a buy-in for the translators. These motivation issues are critical to the success of the crowdsource translation scenarios and this is potentially one of the main reasons why crowdsourcing has not gained widespread success in spite of the major advantageous benefits of crowdsourcing [19].

The txteagle application is building upon the success of Amazon’s Mechanical Turk which allows users to post human intelligent tasks that are usually difficult for computers to complete such as natural language translation [3] . It rewards the users for completing these tasks by way of payment to the users. The txteagle application works off a similar concept by offering users an incentive to use the application by way of MPESA (mobile money) or by offering users free airtime. This incentive is needed to encourage good crowdsourcing as it encourages communities to form which contribute to the group think tank and therefore increase the quality of the product produced by the crowd. Otherwise the situation as discussed in the Facebook example above will happen again by which bad user contributions will be given.

The first prototype of txteagle was distributed in Kenya. The application was distributed in hospitals along a dangerous highway which is home to severe traffic accidents. The blood

supplies in these hospitals were always depleting fairly quickly. The application was used by nurses to update the amount of blood they have available in a hospital to help avoid or signal an upcoming shortage in blood supply [20]. They did this by sending a text to the central blood repository and then they in-turn updated the web site about blood levels. While the nurses had the required mobile equipment they were reluctant to update the service as it was going to cost them air time on their handset. This led to sporadic or non-existent updates of nurses using the system as they were not given an incentive to update the system. In 2007, they updated the system to transfer a small amount of air time to the nurses in exchange for the required information. Once they had given the nurses the incentive to update, it made the system work correctly because the crowd became attracted to the system because it allowed them to participate without the need for paying for the service [20].

The service has now expanded since its successful testing nationwide into Kenya and Rwanda and it is being used in a variety of different tasks by different users such as taxi drivers and local security guards [20].

### **Localisation and Sentiment analysis**

Localization was one of the key aspects to the success of txteagle as there are over 60 distinct languages in Kenya. It is near impossible to incorporate these languages into the service as no translation service currently exists to facilitate this task. The txteagle application had to turn to crowdsourcing to aid in this localization process to help generate a 'phrase book' for the application and to date thanks to the contribution of the crowd more than 15 local languages have been translated using the crowdsourcing model [20].

Sentiment analysis in this application was another key component of its success. Sentiment analysis involves translation that takes the writer's opinion into account when translation of that sentence/paragraph is taking place. There are three ways in which to classify people's opinions which are as positive, negative and neutral. Sentiment analysis is vital in the correct translation of user-based content in the system [20]. A classical example of sentiment analysis would be for translating what bloggers are saying about a brand like Pepsi. Sentiment analysis is also called opinion mining or voice of the customer. This type of analysis is not possible to do via machine translation only via human translation [20,31].

## Quality control

Crowdsourcing translation is not as tightly controlled from the top and opts for a more grassroots approach compared to a professional translation context. This raises an interesting question in relation to how quality control can be imposed in such a decentralised environment. So this leads to three different case types of quality control in crowdsourcing translation which are as follows [19]:

1. The first case is where the customer wants professional translators to revise and approve every translation produced by every member of the crowd. This occurs before it is actually published onto the website of the customer. This can be useful for getting rid of noisy translators who purposely try to upload bad and potentially embarrassing translations onto the customer's website. Example: Facebook
2. The second is where the customer may ask for the translators to go through an initial screening process before they allow them access to translate content. Example: this is used by Translators without borders and Kiva.org.
3. The third approach is where the crowd manage themselves. This type of quality control is carried out using mechanisms such as voting, mutual revision and reputation management of users. Example: this approach is used by Facebook.

There are still many open issues in regards to the cases above between getting a correct balance between getting good quality control on the translation whilst not alienating or segregating members of the crowd. After all if members in the crowd feel like their input is not valuable they will stop contributing to the process of translation. Chris Grams [45] noted that even with crowdsourcing in the successful Wikipedia model there are still some major issues in relation to the quality control aspects that go beyond the question of the users motivations. There are many editors, contributors and readers of Wikipedia that have complained about the current Wikipedia editors who unpredictably change useful content, ban legitimate and sometimes core contributors from the site as well other generally demotivating aspects of which these editors impose on contributors [35,45].

In general professional translation circles have a general assumption that crowdsourcing translation will produce poorer quality output. This may be true depending on the context but it isn't by any means inevitable. There are many different types of situation where crowd source translation can lead to a higher quality of content and this can be done through properly



leveraging of users and giving users a proper incentive so that the wisdom of the crowd effect can be harnessed and used effectively. Currently there are several known principals for making the crowd work collectively smarter than the individuals. The specifics of how these principals can be applied within a given translation contexts is still an open question in the crowdsourcing community [19].

### **Inferring ‘Accuracy’ from Noisy User Responses**

There are major issues in the area of inferring accuracy from multiple error prone users of crowdsourced applications. This issue has been well documented and researched in a variety of different academic papers and in a variety of academic fields. For example Dawid and Skeene 1979, tried to get a more accurate depiction of patient’s medical history based on many different biased medical reports from different clinician physicians [18]. To help to correlate the results Dawid and Skeene introduce a new concept called expectation-maximization (EM) model. This model estimates the bias of these doctors along with the underlying latent variable which in this case is the patient’s record. This approach has been applied across various applications in several different fields such as bio-statistics, image categorization and linguistic annotations. The problem with their application is that they assume all the respondents complete all the tasks that have been assigned. This may not be always the case in crowdsourcing so this approach will have to be altered to fit the crowdsourcing case. The issues can be resolved by an approach discussed by Sheng et al [55] where by the uncertainty of response can be calculated by estimating using a number of samples to achieve a confidence of getting a correct response based on the correlation of responses from users [20,18,55].

### **Critical Success Factors of Mobile Crowdsourcing**

So, what are the factors of creating a successful mobile crowdsourcing application. Ankit (2010) , wrote a research paper based on trying to achieve a better understanding by creating a critical success model. Ankit [54] focuses on mobile crowdsourcing with a particular interest in reviewing the following four projects using this success model which are txtEagle, Ushahidi, Peer Water Exchange and mCollect. The paper also makes a distinction between outsourcing and crowdsourcing. It is noted in the paper that the latter requires and is dependent on the scale of the crowd along with the crowd participation factor. So as discussed previously, the crowd needs incentives to be effective along with also having the motives of crowd being aligned in relation to the goal of the crowdsourcing project. The people in the crowd must use and accept

the crowdsourcing technology that are provided. Ankit models drew upon many other pieces of research to develop the following model [54]:



**Figure 2.5: Ankit Model [54]**

There are five factors in determining the affect of motive alignment of the crowd as illustrated in Figure 2.5. This motive alignment of the crowd is key to the overall success of the crowdsourcing project. The success in the crowdsourcing project is expected to bring in more users and that is why the success and motives is a bi-directional relationship in the model diagram (Figure 2.5).The key elements in model are as follows [54]:

- **Vision and strategy** - the crowdsourcing projects vision and strategy along with the ambition of the crowd will ensure that a crowd of users will take part in the crowdsourcing project.
- **Human Capital** - the more talented and skillful the crowd the less work it takes for the crowd to make meaningful input into the project. This leads towards the potential of a more successful crowdsourcing project.
- **Infrastructure** - the success of mobile crowdsourcing projects rely on the abundance of cheap and reliable mobile communication to ensure that the crowd can take part in the crowdsourcing without having to spend a fortune.
- **Linkages and Trust** - all crowdsourcing projects require time and an information cost from the participants in the crowd. This is why it is very important and critical that the trust factor is developed as well as the addition of linkage/connection factor which can be important in developing this trust along with creating a crowdsourcing community.

- **External environment** - this is described as using the macroeconomic environment which is made up of the following elements which are governance support, business environment, economic environment and risk profiles are all very important in determining the success of the crowdsourcing project.
- **Motive alignment of crowd** - this can be defined as the crowd's ability to relate to the long term goals of the crowdsourcing project and by doing this helping to garner wider crowd participation.

Ankit 2010, applied this model to the four mobile crowdsourcing projects discussed previously and the following table highlights his results:

	txtEagle	Ushahidi	Peer Water Exchange	mcollect
<b>Vision &amp; Strategy</b>	Crowdsourcing simple microtasks; provide supplementary source of income	Crowdsourcing crisis information; ensuring timely crisis response	Clean water access using crowdsourcing; ensuring safe drinking water and sanitation	Enhance export opportunities and trade through crowdsourcing; gather and share price information
<b>Human Capital</b>	Mobile phone skills; Language skills	Mobile phone skills	Mobile phone and internet skills	Mobile phone skills
<b>Linkages &amp; Trust</b>	Started by a MIT doctoral student, supported by Zain, Safaricom mobile service providers	Linkages with Harvard Humanitarian Initiative, OECD and numerous universities in US and Europe	Linkages with Blue Planet Run Foundation, WaterAid, Watershead Organization Trust and Gram Vikas	Linkages with United Nations, World Trade Organization and International Trade Center
<b>Infrastructure</b>	Availability of mobile phones and mobile network	Availability of mobile phones and mobile network	Availability of internet, mobile phones and network	Availability of mobile phones and mobile network
<b>External Environment</b>	Support from the local governments; lack of employment opportunities lead to participation	Proper government support; deteriorating security scenario & occurrence of natural disasters lead to participation	Proper government support; fits in well with the crowd's basic aspirations	Fits in well with the business environment; leads to a better market scenario

**Figure 2.6: Other mobile crowdsourcing projects [54]**

## 2.5 Open issues in mobile translation

### BBC News Mobile web localisation case study

Even with the recent advent of crowdsourcing there are still major challenges in relation to language translation on mobile device such as the need for stable Internet connection on mobile devices and these types of connections can be hard to ascertain in remote areas. The SMS strategy of communicating with the translation server has proven to be less efficient because of the limited message size in comparison to using packets.

Open issue in relation to translation of websites on mobile devices that was discussed by the creators of the BBC World Service website during the Multilingual web conference 2010 in Madrid was the important issue of encoding and font support. They discussed in the conference the general move towards Unicode and the problematic situation of websites using images to solve encoding and font issues. For example mobile displays in countries such as India are still problematic due to the fact that 70% of mobile devices in India cannot display Hindi properly. This has forced the BBC website to display some text based content as images on their website as a temporary solution until the issue of mobile display can be solved. The BBC is currently investigating font download technologies to aid in this field to help Unicode text to become more readable to a wider audience.

These issues have lead to the realisation that the create once publish everywhere factor doesn't exist as different languages have differing issues from encoding, font types and the many issues that need to be addressed when working with right to left scripted languages. Another open issue in relation to web and mobile users is enabling users to input text into websites or editors. The keyboards for certain languages don't necessarily exist and even if they do some don't have all required features.

The BBC International website has over 23 million visitors a week, 10% of which are mobile users and this has lead them to focus on mobile support for their website. This introduction and increasing use of mobile devices has lead to reintroducing many of the old issues related to font support. The problem usually is that the font is not supported or OpenType (format for scalable computer fonts) features are not available. This issue in relation to font has lead to over 70% of mobile devices not rendering text on web pages correctly thus leading to a loss in translation [39,68].

## **2.6 Technology Research**

There are many crucial aspects that need to be researched in relation to the technologies that are going to be used in the creation of the prototype in this dissertation such as:

- Localisation management tools
- Semantic Web
- Cloud technologies
- Application technologies
- Mobile technologies

### **Localisation management tools (SOLAS)**

SOLAS [42] is a collaborative project by CNGL which has collaborators in Irish universities such as University of Limerick (UL), Trinity College Dublin (TCD) and Dublin City University (DCU). Companies large and small have localisation needs as they trade internationally in addition to the corporate context there are many different non for profit organisations that strive to provide localised content for under resourced communities. The localisation process though is a long difficult process but thankfully nowadays there are localisation management tools already available to a project manager to help ease the process of localisation. An organisation in particular small organisation may not have the resources or the expertise to use these available localization management tools.

SOLAS [42] however has taken a fresh approach for tackling these issues by simplifying the localisation/translation process by making an agile, flexible and lightweight platform for localisation. The SOLAS application allows for such a highly configurable localisation process because SOLAS is made up of group of components and these components are inter changeable dependent on the type of localisation job. These components can be created and simply plugged-in to the SOLAS platform.

LocConnect component allows the project manager to create, track and manage translation projects. It makes the creation of the new translation jobs for a document easy for the project manager and the work flow will be built up dependent on the jobs requirements. Once a job and the relevant translation document have been upload they go through other components in SOLAS [42] depending on the jobs requirements this is done by the workflow recommender .To make sure a translation job can be successful the source content needs to be of a good enough quality. This is where localisation knowledge repository component comes into use because with the help of this component, the source content can be improved by editing using recommended changes to the content.

To help with the translation of documents it makes sense to make use of existing translations rather than having all the content translated again. This is done via XLIFF file format using the XLIFF Phoenix component. The XLIFF file format is an extension of XML which was created to standardize localization. In 2002, OASIS standardized the XLIFF standard [43]. The XLIFF file format is used in XLIFF Phoenix component to help store and share localization data.

Machine Translation mapper in SOLAS is the machine translation component of the application and this step is taken care of automatically by SOLAS in the work flow process. The application also has a crowd sourcing rating component. This rating component in essence asks each member of the community to vote on translations related to a relevant job. This component is optional in the creation of new jobs. [42]



```
<?xml version="1.0" encoding="windows-1252" ?>
<xliiff version="1.1" xmlns="urn:oasis:names:tc:xliiff:document:1.1">
  <file source-language="en" original="localization.localizationBundle"
        datatype="x-oracle-adf">
    <body>
      <trans-unit id="localization.Caldroptest.Starttime_LABEL">
        <source>Startzeit</source>
        <target/>
      </trans-unit>
      <trans-unit id="localization.Caldroptest.Endtime_LABEL">
        <source>Endzeit</source>
        <target/>
      </trans-unit>
    </body>
  </file>
</xliiff>
```

**Figure 2.7: XLIFF example[44]**

Many of design aspects in SOLAS could be useful in the creation of this projects prototype because it is currently the state of the art in next generation localisation. Aspects in the design of SOLAS [42] that have informed the design of the prototype are:

- XLIFF usage for storage and sharing of data and job creation mechanisms,
- The use of Machine translation with post-editing and
- The use of crowdsourced rating of other contributors content in the community.

### **Semantic Web**

There are few aspects in the design of SOLAS [42] that can be improved on such as the fact that all localisation data in the application is stored via a relational database management system (RDBMS). A relational database management system can be defined as database management system in which all data is stored in tables along with the relationships among the data. This data stored in the tables can be accessed and reassembled in many different ways without affecting the table's structures. A better possible way to store this linguistic data in such a localisation application is via a triple store which is like RDBMS in that both are general purpose data storage engines.

The trade-offs between both types of storage engines arise in terms of the flexibility, performance, cost, robustness etc. RDBMS is suitable for online transaction processing where there needs to be fast transactional updates such as in e-commerce applications. In the RDBMS scenario the data scheme is usually small and all the data is known in advance so that the data tables can be made to hold the data. RDBMS primarily focuses on performance, resilience and availability to a large system of tools and services. This would be useful for creating the storage for user data, job data, comments and ratings in the prototype.

The triple store framework though makes use of semantic web mechanisms such RDF data as well as giving mechanisms for persistent storage and access to RDF graphs . The Semantic web can be defined as a “web of data” and it was a phrase coined by Tim-Berners Lee the inventor of the World Wide Web and the current head of the world wide web consortium (W3C). Tim-Berners Lee defined the semantic web as *“a web of data that can be processed directly and indirectly by machines”* [10].

So basically, semantic web tools such as RDF are an extension of the World Wide Web such that they enable web users to share their content beyond the confines of a single web site or application [51]. This move to semantic data reflects the natural societal change in every-bodies day-to-day use of the web (example: Facebook, microformats, Wikipedia etc).

*“I have a dream for the web [in which computers] become capable of analyzing all the data on the Web – the content, links and transactions between people and computers. A ‘Semantic Web’, which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The ‘intelligent agents’ people have touted for ages will finally materialize.” – Tim Berners-Lee, 1999*

Resource Description Framework (RDF) is a language that is used for representing information about resources on the world wide web. RDF was intended to represent meta data about resources on the web such as author, title, modification data of site etc. Identifying this RDF data is done through the use of Uniform Resource Identifiers (URIs) along with using simple property values to describe the resources. This enables RDF to represent simple statements about resources as a *graph* of nodes and arcs representing the resources and their properties and values [72]. To take an example *“there is a Person identified by*

*http://www.w3.org/People/EM/contact#me*, whose name is Eric Miller, whose email address is *em@w3.org*, and whose title is Dr." the RDF graph results can be seen in Figure 2.8:



**Figure 2.8: An example RDF graph [72]**

RDF triple stores are used in scenarios where the data schema is less defined. This can be useful where there is a wide range of different types of data to be stored as well as an evolving set of relationships between the data. This can be useful for building up linked relationships between translation jobs which use similar languages, are in the same job categories etc.

This makes the environment of the triple store flexible and schema-free and makes dealing with this type of evolving data easy and convenient. The triple stores can be useful as well for integrating data from a wide range of sources but don't need to force the data into a single normalised schema. The triple store is an agile backing store that can be used for fast prototyping and experimentation and it does not require a lot of creation of tables or linking relationships of columns [50].

Those are not the only advantages of using triple stores over RDBMS for localisation of data there are many other advantages particular in the modern business of industrial localisation. This is because the face of industrial localisation has changed from periodic translation of a large body of source content to small (long-tail) and diverse translation of content which needs to be translated in a fast and user demand driven fashion. Small and medium sized enterprises can



compete with these larger companies in this new localisation market by allowing the reuse, capturing and sharing of multilingual data that has been created in the process of localisation by using the data stored in the triple store. This multilingual data that has been created becomes essential for process iteration on new data and to help in the training of domain-specific data driven language technologies such as statistical machine translation (SMT) [32].

The language resources held in the triple stores can be put to effective use in long-tail localization also such that there is low-cost access to good quality translation material which can be reused again [32]. So the solution can take the form of linked data on the web such that good quality assured localisation material can be ascertained via the triple stores RDF data. The semantic solution compared to RDBMS is also more suitable as it can inherently handle multilingual data with element-level tagging in the RDF data. This can be used regardless of the required language and this can be useful for addressing the needs of many small to medium sized enterprises. This research by Lewis (2011), is informed by work taking place through EU funding in the ICT domain in projects such as [32]:

- language resource networks (FLaReNet project),
- visions of multi-lingual web content (Multi-Lingual Web project)
- development of language/localisation technology roadmaps
- federated language resource repositories (META-NET project)

### **Cloud technologies**

The term cloud computing is the provision of virtual computational resources that can be accessed on demand via the use of a computer network. These resources can maintain and manage themselves across any small to large scale server cluster with resources that can include the following services:

- Application hosting
- Database services
- File services, Email etc....

The cloud computing model takes all the heavy computational work off the local devices and concentrates these resources on the cloud. The cloud offers other advantages such as the ability to host and manage web based applications without having to worry about the tedious technical details so that the business logic of the crowdsourcing application can be focused on. This leads

to an application that does not need to be installed on the user's handset but instead users can access the application via the cloud using a web browser. There are many successful examples of cloud and crowdsourcing applications such as Wikipedia, Amazon Mechanical Turk etc [3]. The cloud therefore naturally lends itself to crowdsourced based applications in the way it allows for the following [15,60]:

- **On demand access** - for participants in the crowd.
- **Easy replication and fault tolerance** - in the face of failure which allows for vital work of the crowd to be secure and safe.
- **Extremely scalable** - for allowing for fluctuation in increasing/decreasing the size of the crowd.
- **Accessibility** - the application will be accessible from anywhere once the user has a Wi-Fi or 3G connection to the web.
- **Application updates** - updates to the application can occur naturally and by simply updating the instance of the application on the cloud. This means users will not have to download or update their application through the use of patches as the application will be accessed and managed via the cloud.

### **Application technologies and Mobile technologies**

There are two main options in the design of the prototype which will be smart phone based and these options are a mobile application or a web browser based application. The best option that has the potential to reduce coding effort is clearly a web browser based application. This is because for example iPhone applications are created using Objective-C [7] and Android applications are created using Java [4]. So the best option, for creating a portable application would be a web based application which would be able to work on many different platforms from iPhone to Android to Blackberry to Windows mobile.

Typically a web development environment can be split into a multitude of areas but a basic web development hierarchy can be split into two distinct areas which are client-side and server-side. There are lots of differing client-side coding options such as Silverlight, Flash, HTML and CSS, Ajax and JavaScript. Since the prototype is going to be an ascetically simple mobile localisation application there will be no need for Silverlight and Flash. This is also advantageous because Flash and Silverlight capabilities on smart mobile phones could detract from users experience as Flash and Silverlight causes extra burden on the client's side connection. So it

would be best on the client-side to use just HTML and CSS along with JavaScript and Ajax where necessary.

There are also a lot of options for the server-side coding such as ASP, CSP, JSP, PHP, Ruby, and Python etc. The main thing to take into consideration when choosing which server-side coding language to use is what types of libraries will be used on the server side. On the server side for example there will be a lot of semantic web tools used such as a triple store. The most popular triple store libraries are coded using Java such as Sesame and Jena. There will also be some RDBMs interactions which can be handle via Java libraries such as Java Database Connectivity (JDBC). This makes JSP the best option for the server-side coding. This research information is very advantageous in relation to the creation of the application as it helps to highlight a scalable, extensible and highlight portable solution using JSP. An important aspect to consider in the development and testing process is to test the application on all the major mobile web browsers. The major mobile web browsers at the moment are as follows:

<b>Mobile Web-Browsers</b>	<b>Supported Mobile OS</b>
Android Browser	Android
Blackberry Browser	Blackberry
Internet Explorer Mobile	Windows Mobile
Safari	iPhone
Opera Mobile	Android, Symbian and Windows Mobile
Firefox for mobile	Android

**Figure 2.9: Popular Mobile Web browsers [61]**

## **2.7 Conclusion**

Since the inception of mobile phones in recent years humans have been empowered to connect and communicate with one another in real time along with allowing users to conduct real-time peer-to-peer transactions . It has been noticed that it is troublesome for organisations to hire local users in remote areas in places such as Africa [20] to complete simple tasks that can help to contribute towards crowdsourcing projects. It is the view of this author, that the introduction

of incentives for these users may encourage them to contribute towards the crowdsourcing initiative whilst also having a profound and positive effect towards these users' lives. Aspects to keep in mind during the design of the application is to make sure the motives of the crowd are aligned and that users in the crowd are managed in a fashion that doesn't alienate certain users of the crowd.

There are a few issues though in the creation of these crowdsourcing projects such as issues with encoding of text which was highlighted in the Madrid conference by the BBC World Service. The BBC World Service [39] came to the conclusion that the create once publish everywhere factor doesn't exist as different languages have different encoding issues, font types and a whole host of different issues when working with right to left scripted languages. Another open issue is the whole issue that certain keyboards do not exist for certain languages and even if the keyboards do exist all of the required features may not be available. To fix this issue will require expertise from a person who speaks the language so that a suitable alternative can be created for users of these types of languages such as the creation of a feature built into the web application [39].

The current localisation applications in development such as SOLAS [42] don't have a mobile version and on top of this, all of the localized content is being stored in relational databases instead of an RDF triple store. The use of an RDF triple store could put the language resources stored in the triple store to effective use in long-tail localisation such that there is low-cost access to good quality material which was used in prior localised tasks which can be found in the linked data. So the solution can take the form of linked data on the web such that good quality assured localisation material can be reused and ascertained via the triple stores RDF data. The use of RDF will also be useful in building up a graph of relationships between translation jobs which use similar languages, are in the same job categories etc.

To conclude this state of the art research, it has noted that a cloud based crowdsourcing solution would be best. This is because it would allow any user with a smart mobile phone on demand access anywhere in the world whilst also allowing a highly scalable and fault tolerant solution for the creation for the application. In this vein, it was decided that web browser solution would reach a wider audience of smart phone users in comparison to a application based solution written in Java [4] working only on Android or Objective-C [7] solution working only on iPhone. A web browser solution written using JavaScript, Ajax, CSS and HTML with JSP is the

best solution as it will allow all mobile web browsers to support the application from iPhone to Android to Blackberry to Windows mobile. The use of JSP would conveniently fit in with the use of semantic web tools for storing localisation content because most of the triple stores libraries are written in Java.

# Chapter 3

## Design

This chapter's purpose is to make informed design decisions for the mobile crowdsourcing translation prototype based on the review of the current state of the art. This design chapter will help to ascertain the principal requirements of the proposed prototype along with discussing and analysing the design choices. Through this discussion of these design choices a more in-depth understanding of the system can be achieved.

### **3.1 Functional Requirements**

Functional requirements are used to capture the required behaviour of a system. This system behaviour can be described as the services, tasks or functions that a system may have to perform in its environment during its lifetime [8]. The state of the art review has helped in obtaining these functional requirements which are:

#### **3.1.1 Sociability**

To help users to contribute to the crowd users should feel that their input into the community is important. To help provide for this social aspect users contribution to the localisation process should be clearly and highly visible in all of the user's contributions from creating localisation jobs to offering possible translations. To reward the users for usage there will be a reputation board on the website which will contain a reputation score which is based on ratings from user's contributions to the website. To add in extra sociability measures all the users in the prototype will have user pages which will give a brief showing of the user's reputation score, the user's interests and the jobs that the user has created. These user pages are accessible by clicking on users names in the application. Other sociability aspects available in the application will be commenting and rating of other users contributions.

#### **3.1.2 Usability**

The main aim in the usability of the application is that the prototype is easy and intuitive to use as well as being enjoyable to use on a mobile handset and desktop environment. This is achieved in the application through the use of human computer interaction (HCI) techniques such as prototyping, storyboarding and user testing.

## **3.2 Design Choices**

### **3.2.1 Project Scoping**

The state of the art has allowed in narrowing down the scope of the crowdsourcing application to suit the mobile domain.

In section 2.5, BBC World Service website raised the important issue of encoding and font support on mobile devices during the Multilingual web conference 2010 in Madrid. They highlighted an example about mobile display in India which are still problematic due to the fact that 70% of mobile devices in India cannot display Hindi properly (right to left language) [39,68]. This issue has helped to highlight the many issues there are with working with right to left scripted languages on mobile handsets. For this reason, all right to left scripted languages will be deemed beyond the scope of this project and therefore languages handled in this prototype will be left to right and will be primarily European based such as English, Spanish, German, Italian, Romanian, Portuguese and Dutch.

### **3.2.2 Data Storage**

In the state of the art in section 2.6.2, the use of semantic web was highlighted in the form of triple stores. The triple store will be used for storing localisation data (source and translated content) along with references to the job it is related in areas such as job number, word count, jobs creator, job creation date etc. At the present time there are a lot of triple stores on the market but the two most popular are Jena and Sesame [11]. This section will evaluate these triple stores for suitability and performance to help in the selection of the most suitable native triple store.

Research done by Bizer et al helps to evaluate this situation in relation to scalability and performance. The performance of Jena for querying data from the triple store is slower than Sesame such that an evaluation up to 100 million data set queries show that Sesame has faster response times than a Jena. The negative aspect of Sesame is that the loading and evaluation of 1 million data set is much faster using Jena . The load churn in the application will not exceed these levels so loading of data should not be a major factor. The querying of data on the other hand should be fast and efficient and in particular where there is a large body of users using the application. Therefore taking these performance aspects based on Bizer et al research into consideration it is clear that Sesame seems like a better all-round solution for data sizes (<100 Million) assuming that there is infrequent loads/low data churn [12,11].

<b>Tasks</b>	<b>Jena</b>	<b>Sesame</b>
<b>Load 1 million triples of evaluation data:</b>	2 minutes 59 seconds	<b>49 Seconds</b>
<b>Simple Query Mixes per Hour for 1 Million triples:</b>	<b>38,727</b>	15,842
<b>Simple Query Mixes per Hour for 100 Million triples:</b>	<b>3,116</b>	459
<b>Complex Query Mixes per Hour for 1 Million triples:</b>	<b>18,094</b>	4,450
<b>Complex Query Mixes per Hour for 100 Million triples:</b>	<b>254</b>	81

**Figure 3.1: Sesame Vs. Jena [12]**

All other data in the application related to users account, job details, ratings and comments are going to be stored in a RDBMS database. There are a variety of different RDBMS databases such as MySQL, Oracle and PostgreSQL. The usage of MySQL has been deemed best for the development of the prototype of this application.

### **3.3 Key Components**

#### **3.3.1 User Pages**

Users of the application will be given a user page once they have registered with the application. When the user registers an account they will be asked for their name, email, a unique user name for their account and a password. Once a user has registered an account they can log in, contribute to the community and they get their own user page. User pages can be accessed via hyperlinks which are available whenever a user name is displayed in the application.



All user pages can be personalized by the use of a small biography along with a list of their interests. This type of personalization helps to give extra scope for the functional requirement of sociability in the application because it allows user to learn more about the other members in the community whilst also helping to strengthen the ties between members in the community. Some crowdsourcing websites such as Wikipedia advocate this personalization of user pages and they have proved to be very successful among their more prolific contributors by allowing them to express their individuality on the website thus creating a stronger sociability on the website.

The main aim of the user page other than promoting sociability on the website is allow users to view their jobs, other user's jobs, private jobs which other users have invited them to take part in and old jobs that they have created but have reached their completion date. This gives the users a nice area which users can conveniently manage the jobs that they are involved with in the application as well as viewing their old job and current reputation score.

### 3.3.2 Job Life Cycle

There are three main aspects to the life cycle for a job in the application which are as follows:



Figure 3.2: Job Life Cycle

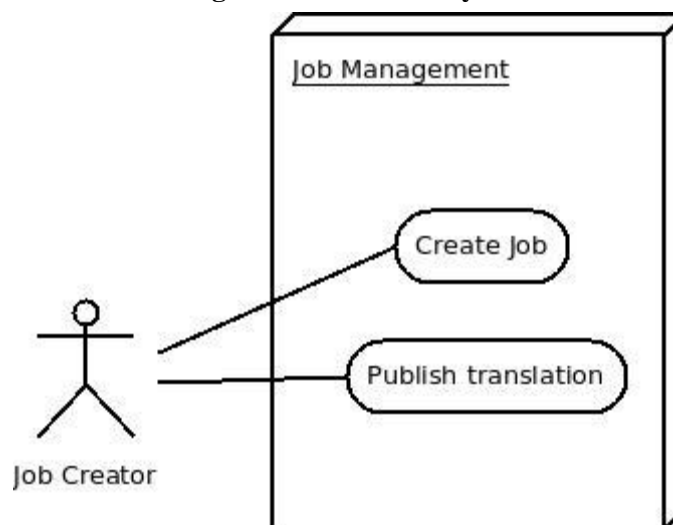
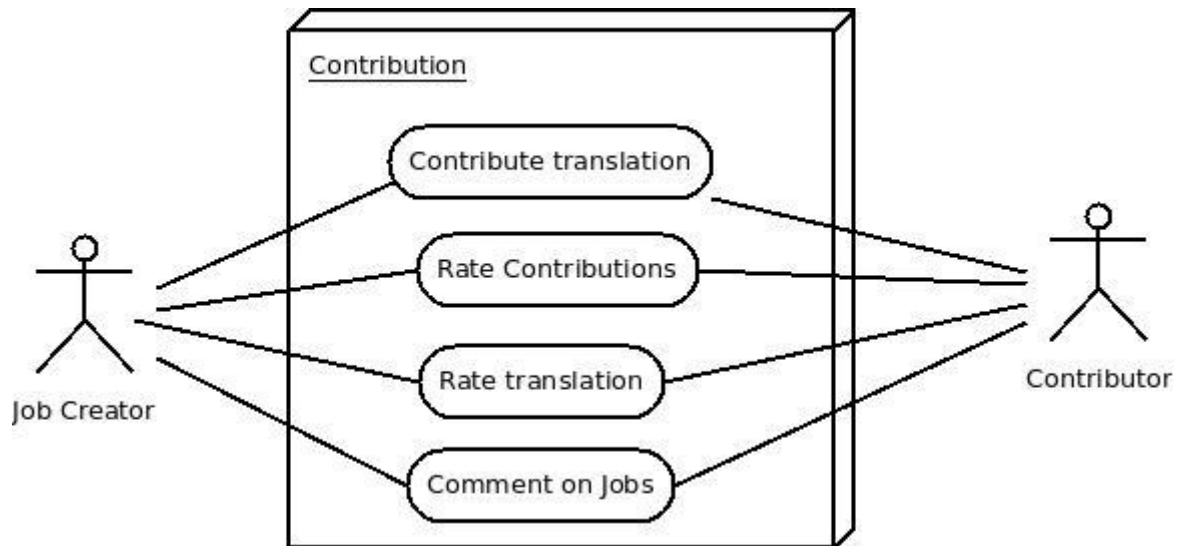


Figure 3.3: Job Management

## Creation

The creation is the first step in the life cycle of a job in the application. The creation of the job should be simply done in that it should just consist of the name of the job, a brief description, a start date, a end date, the language to translate from, the language to translate to, category, select job distribution (public or private), select people to send job to and the source content to be translated. If the job is public it goes to all other users' inbox and is available to see in the portal (home) page. If the job is private, the job will only be accessible and visible to the creator of the job along with users the creators has invited to the job. Once the job is created the source content to be translated is stored in the triple store using the RDFLogger component (section 3.3.7) whilst all other data relating to the job is stored in the MySQL database.

## Contribution



**Figure 3.4: Job Contribution**

Contributions to the job can be done by the job creator and other contributors in the community who have been granted access to the job. This JSP page that takes user contributions must be very well designed and very usable as it will have to be usable on small screens such as on mobile handsets. The source content that must be translated will be on top followed by the list of contributed content from users which has been retrieved from the triple store along with the like and dislike ratings given by users. The user can enter their own translation or they can use a machine translation to translate the source content to the required language which is indicated on top of the page and on the top of the area in which user enters their contribution translation .

There will also be an area where the source content that needs to be translated is broken up into its sentence so that users can clearly see the sentences that need to be translated. Once the user enters their translation it should be uploaded to the triple store using the RDFLogger.

### **Publishing**

The publishing page will have the job of publishing the results of the localization job and the publication will be decided by the creator of the job (as indicated in Figure 3.3) when the end date has been reached on the job. The publication page will be laid out with the source content on top followed underneath by user contributions and the rating that the translation was given. The creator then picks the best translation from the bunch based on either his own point of view or he can base his decision on the crowd and go with the highest rated translation or the user may not want to publish because none of the translations were good enough. If the job is published it is then uploaded by the RDFLogger to the triple store. This is the final step in the jobs life cycle.

#### **3.3.3 Deliberations**

##### **Comments**

There are a lot of options for discussing jobs and user's contributions such as forums or Wiki but none of these options will integrate seamlessly into the contribution page. The best way to integrate user comments into the application is through the use of user comments similar to YouTube or Facebook where users can comment and respond to other user's comments directly. This feature will be hosted at the bottom of the contribution page and users can add comments by simply clicking comment to add a new comment. There will also be a feature where users can vote on other users comments by either clicking "like" which will increase that users reputation score or by clicking dislike which will lessen a user's reputation score in the community.

##### **Rating**

Similar to user comments there were a lot of options for creating a rating system for the application such as a star rating between 0 and 5 or the use of "like" and "dislike" rating system. The latter option of a "like" or "dislike" rating system will be used in the application and will be similar to that seen on YouTube [42]. This option will fit the more compact restricted area of the mobile handset compared to users selecting a star rating. This rating mechanism will be available for rating user translations on a job along with being available for rating user comments. The "like" mechanism will increase users reputation score in the application and

“dislike” will decrease user’s reputation score in the application. All the user rating scores are public and are published alongside the translation or the users comment. The members of the community can use this rating system any way they want as it is ultimately up to the community to form their own rules according to how the rating system is used.

These mechanisms of rating and commenting ultimately add another dimension of sociability onto the application by letting users in the community voice their opinions on translation jobs by means of commenting and rating. These features ultimately will help strengthen community participation and feedback in the community.

### **3.3.4 Translation**

Some pieces of the content that need to be translated for a particular job could be long and this length could deter other users from contributing to the task or it could slow down contributions of users to other tasks because the current volume of data in the task that they are working on is large . However, despite the large volume of task content to be translated the need for localisation of the content still exists for the job creator. To make this process more accessible to work on for users the use of machine translation could be very beneficial. It could be very beneficial because machine translation engines help to translate content instantly but the output of the quality of the translation can be of a poorer quality compared to human translation.

The best design solution for the translation process of large, medium or small content would be a hybrid of both machine and human translation. This solution would work by having the machine do an initial translation on the content and then the machine translation would be revised and post-edited by a motivated user in the crowd thus producing a fast high quality translation of the content. This kind of work of post-editing by translators can be increasingly found these days where by translators are engaging in quality control and terminological roles instead of straight forward translating of content [29].

There are a lot of different translation APIs and libraries available the most popular are Google Language API, myGengo, babelfish and Microsoft translator API. Initially the machine translation component was going to be designed using the Google Language API but on May 26, 2011 Google announced the official deprecation of the API and scheduled shut down of the API in December. To help fill this void a Java based library called Microsoft-translator-java-api was created which is a Java wrapper around the Microsoft Translator API [23]. This Java library

mimics the code structure, name conversion, usage patterns and functionality to mimic the widely known Google Language API. The only requirement for using the Java library is attaining a Bing developer API key from Microsoft which is easily done via the Microsoft website.

The languages that will be available in the machine translation component in the application are English, Spanish, German, Italian, Romanian, Portuguese and Dutch which is in keeping with the scope of the project outlined in section 3.2.1. The design will consist of two text boxes one will contain the data that needs to be translated and the other box contains the translated data. Both the source content and translated content can be edited by the user to help achieve the best translation possible. The user will also have the option to ignore the machine translation and do their own translation as they see fit. These translations once completed can be uploaded to the triple store to the associated job by the user.

### **3.3.5 Keyboards**

As the application expands it will have to take on board potentially rarer languages which may not have keyboard support on the major smart phone handsets such as rare local dialects in Africa. As discussed in section 2.7, this issue of keyboard support could be potentially fixed by two options which are a web browser plug-in or feature a built in keyboard for the languages in the application. The best option would be to build in keyboard for the application so that custom keyboard support would be available on all web browsers that run the application.

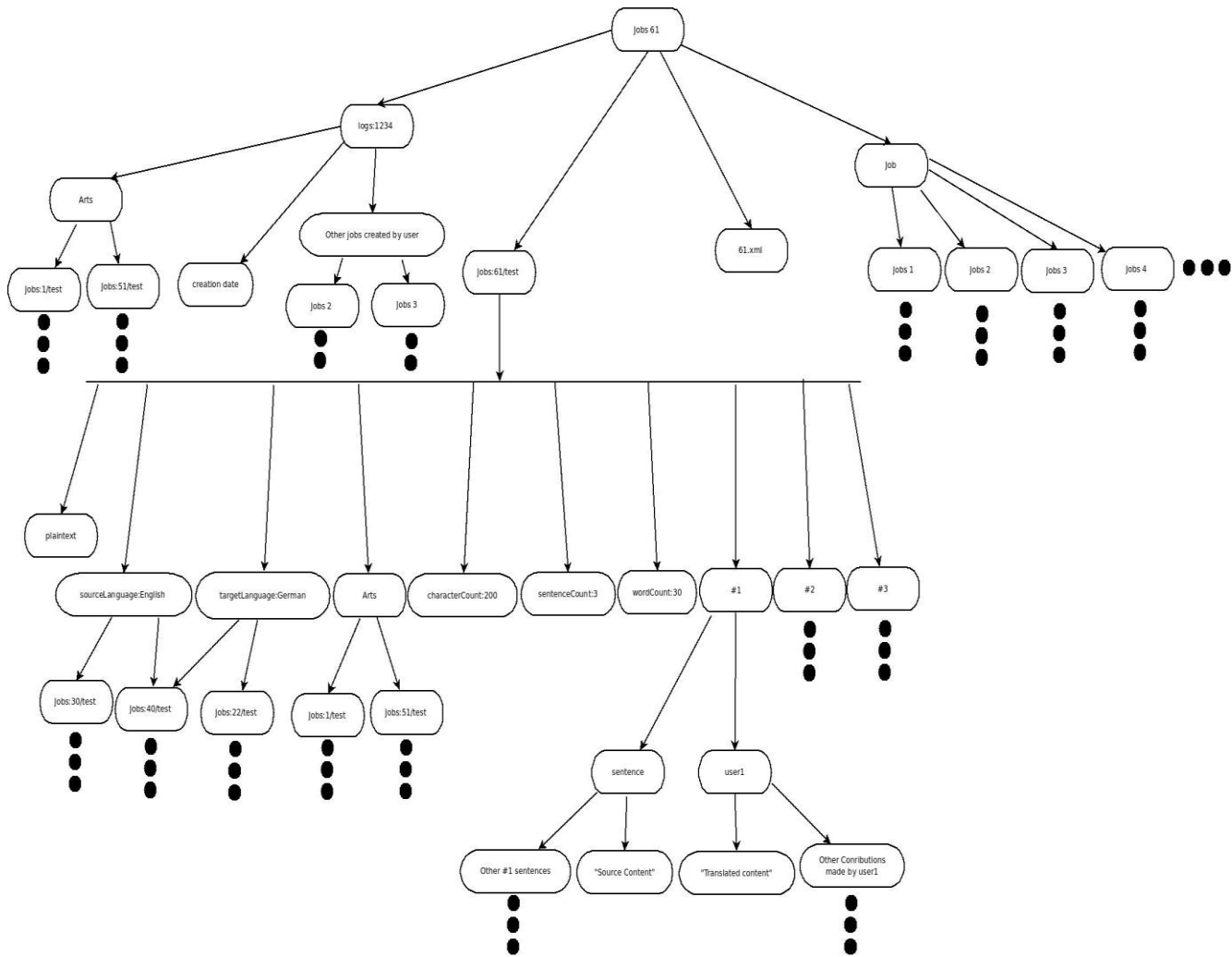
To keep things simple for the initial prototype keyboard, support will be only available for the languages highlighted within the project scope which is discussed in section 3.2.1. The virtual keyboard design will be defined by HTML and CSS while JavaScript will be used to control the behaviour of the keyboard.

### **3.3.6 Data Model**

#### **RDF**

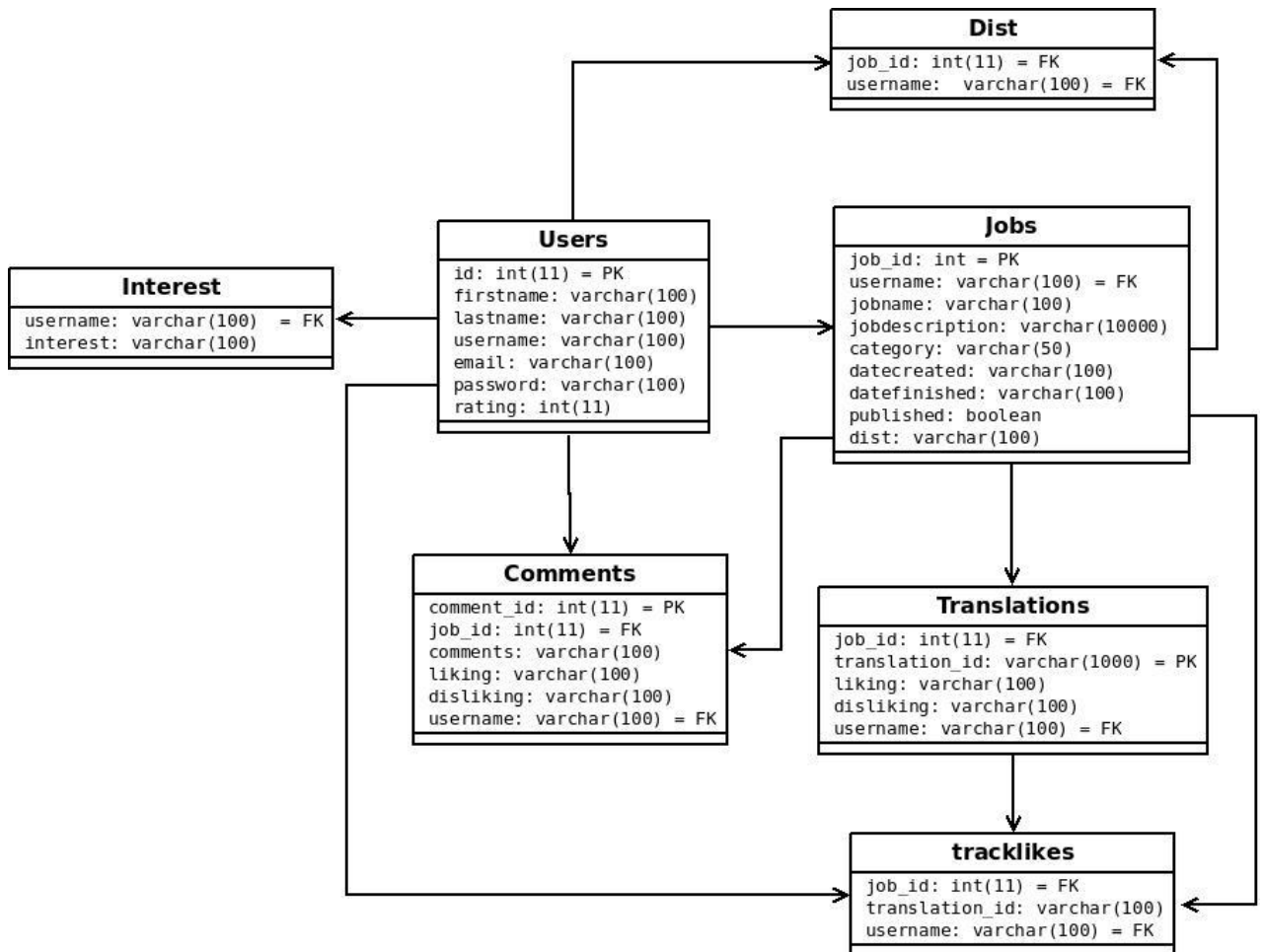
As discussed in section 2.6 in the State of the Art data in the application will be stored in two locations which are a triple store and a MySQL database. All language translation data will be stored in the RDF triple store and all other data relating user's accounts, job details, ratings and comments will be stored in a MySQL database. Figure 3.5 backs up the reason discussed in State of the Art that storing data in the RDF database would be better than RDBMS because relationships between the data can be more easily and organically built up over time. These

relationships between the data can be seen in many areas of Figure 3.5 with the graph example where job 61 in the example is in the Arts category and therefore can be linked to all other jobs in the art category including the localisation and information data from those jobs. Similarly all English translate to jobs and German translate from jobs are linked to job 61 etc. This ever expanding and infinite looping data structure would not be possible to map in the MySQL structure were all jobs are interlinked together. This interlinking of relationships between data will be advantageous in helping in long-tail localization jobs and in aiding in the teaching of statistical machine translations as discussed in the state of the art. A sample of the RDF structure that was used in the prototype can be seen in the Figure 3.5 below:



**Figure 3.5: Sample RDF graph data from Prototype**

## MySQL



**Figure 3.6: MySQL data model**

MySQL 5.0 was used for the back end for handling data such as user's accounts, job details, ratings and comments. There are a total of seven tables used in the application which are as follows:

1. **Jobs** - this table is used to store data about the jobs such as job identification number, job name, user that created the job, a job description, the category of the job, the distribution of the job (public or private), the date the job was created, the date the job is to be finished and has the job been published yet.
2. **Users** - this table is used to store data about user accounts such as first name, last name, user name, email address, password and the rating of the users.

3. **Tracklikes** - this table is used to store the ratings and the contributors of translations in the application and it holds data such as the job identification number, the like rating of the translation, the dislike rating of the translation and the user that contributed the translation.
4. **Distribution** - this table is used to highlight the users that have been distributed a particular private job it contains data such as the job identification number and the user name of the user which the job has to be distributed to.
5. **Comments** - this table is used to hold all the users comments on jobs in the prototype. It contains data such as the job identification number, the comment, the like rating of the comment and the dislike rating of the comment.
6. **Interest** - this table is used to store the details of users interest in the application. It contains data such as the user name of the user followed by a list of the users interests.
7. **Translations** - this table stores the user translation contribution details such as the job identification number which the contribution is linked too in the triple store, the user name of the person who gave the contribution and the like and dislike rating score.

The creation commands of these six tables can be found in Appendix A.

### 3.3.7 RDFLogger

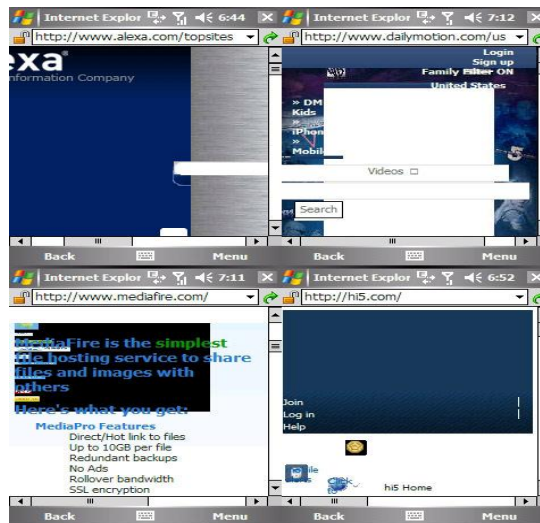
This component was built by Sebastian Moliness as part of a CNGL funded project called CnglRdf. The RDFLogger is basically a servlet which can change an XLIFF document into RDF statements. This conversion from an XLIFF document to RDF statements (benefits of RDF discussed in section 3.3.6) is done through XSL transformation (XSLT) which processes the XLIFF document to create Notation3 (N3) statements. N3 is an alternative way of modelling RDF's XML syntax so that it is more compact and human-readable as well as allowing greater expressiveness in the data through the support of RDF-based rules [69].

Once these RDF statements have been created they then can be merged into the Sesame triple store. The RDFLogger then produces extra RDF statements to track provenance information such as unique "log" URI. This RDFLogger component will be essential because it will be used to enter new job data that needs to be localised into the triple store along updating the triple store with users translation for the jobs.



### 3.3.8 Mobile design

The major aspect to take into consideration when assigning styles to your website for mobile devices is to keep the style of the website very simple. The style should be kept very simple because mobile devices are slower than their computer counterparts, so avoiding the use of lots of images and graphics is advantageous towards the usability and loading times of the application [70]. The other important aspect to take into consideration is the various screen sizes and resolutions in mobile devices. Therefore setting of page elements by fixed pixel widths and heights is a bad idea as the web page may not display properly due to the screen size [47]. The results of badly formatted mobile web pages can be seen in Figure 3.7.



**Figure 3.7: Examples of badly formatted websites [46]**

The solution to this problem is to simply scale the page elements based on the screen size this will format the web page on any mobile device correctly. The results of properly formatted mobile web pages can be seen in Figure 3.8.



**Figure 3.8: Examples of correctly formatted websites [46]**

This scaling of the web pages correctly can be achieved through many different options such as two different CSS pages one for desktop screens while the other is for mobile screens or the use of viewport META tag in the HTML to indicate what scale to render the page for mobile devices. The use of META tags and simplistic minimalistic design will lead to a design in the web application which will have less coding than the two CSS pages option. It will also give a more portable solution which will work and format properly along with giving the application a similar look and feel on both mobile and computer browsers without the extra coding effort. The viewport META tag is used to control the dimensions and scaling of the viewport window in the mobile web browsers [30]. The viewport META tag is compatible in all the major mobile web browsers such as Safari Mobile, Android browser, BlackBerry browsers and Opera browser.

```
<meta name="viewport" content="width=device-width, initial-scale=1.0, maximum-scale=2.0,
user-scalable=yes" />
```

**Figure 3.9: Viewport META TAG [49]**

The meanings of the elements in the viewport META tag are as follows [49]:

**width**

The majority of smart phone browsers will scale to a default wide viewport which will show an overview of the whole web page. To get a closer and more usable view of the web page the viewport must be set so that the default viewport value is not used. This can be done easily by the use of the width=device-width.

**initial-scale**

The initial scale is the initial zoom in or zoom out scale of the web page when it loads onto the mobile web browser. The best and recommended initial scale to set the web page is 1 because the text size will be at the optimal size for users of the web pages.

**maximum-scale**

This maximum scale is used to limit how far users can zoom into the web page. The best and recommended limit is 2 so that users can zoom into the web page in case they want a zoom in to view some text.

**user-scalable**

This is basically used to specify whether users can or cannot zoom in or zoom out on a web page. It is always better to have this option available in case users want to get a closer view on text on the web page.

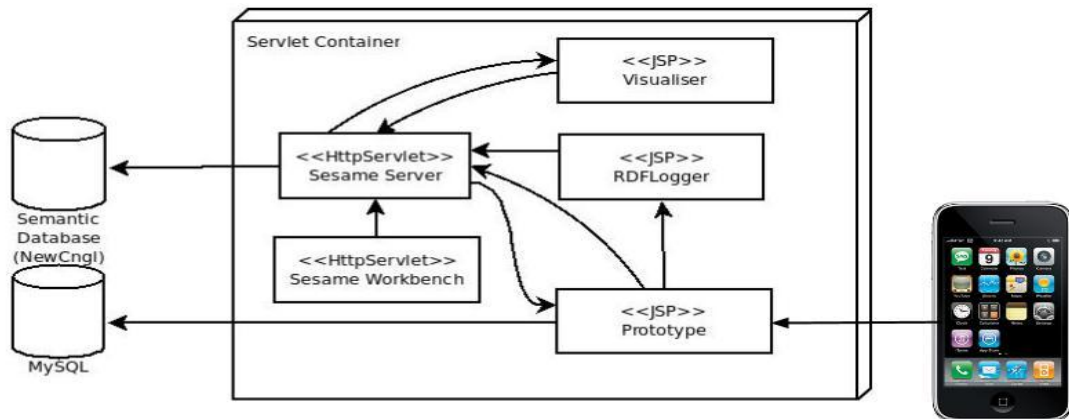
# Chapter 4

## Implementation

This chapter will justify the implementation decisions made in the project along with the practical issues that arose during the implementation of the mobile crowdsourcing localisation prototype. This chapter will also discuss the key features of the prototype.

### 4.1 Deployment

The deployment of the application is illustrated below in Figure 4.1. All the components in the diagram are servlets and these components can either be deployed on one JSP servlet container or the components can be deployed on distinct servlet containers. The deployment of this prototype was done onto one JSP servlet container. The JSP servlet container used in this project was Apache Tomcat [6].



**Figure 4.1:Deployment**

The **Prototype** component will be the subject of most of the discussion because it was the component created during the development of this project. The **RDFLogger** component was developed by CNGl and as discussed in section 3.3.7 this component is used to change an XLIFF document into RDF statements for storage in the triple store. The **Sesame servlets** are an open source Java framework for storing, querying and reasoning with RDF and RDF schema [52]. These servlets will now be discussed in more detail in the sections below.

#### 4.1.1 Sesame Server

The Sesame server component is used as a server with which the applications (Visualiser, RDFLogger and Prototype component) can communication with over HTTP. The main component of the sesame server (and framework) is the repository which is a storage container for RDF data. All the operations such as creating a repository, querying repository and updating repository data in Sesame is done with respect to the repository. There are a number of different types of repositories that can be used in Sesame such as In-Memory store , Native Java store or a Remote RDF store. The particular deployment used in this project was an In-Memory store.

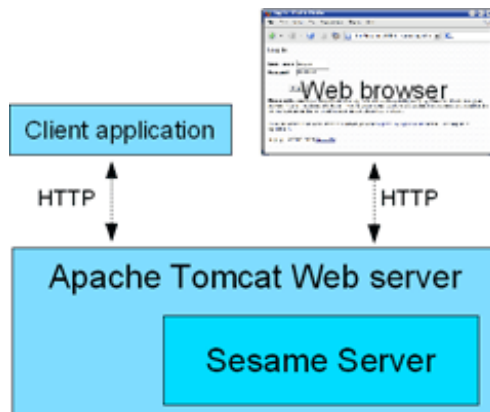


Figure 4.2: Sesame server [52]

#### 4.1.2 Sesame Workbench

The workbench component is used as a functional administration interface which can be easily accessed via any web browser. The interface can be accessed by adding openrdf-workbench onto the URL of the server on which you are hosting Sesame for example <http://hostingserver.com/openrdf-workbench>. The workbench allows for administration of the semantic data by allowing administrators to create new data repositories (graphs), adding RDF data along with exploring and querying data in the repository. The data repository created for this project was a In-Memory store called NewCNGL. The In-Memory store was chosen because it is the fastest type of repository since the entire RDF graph is kept in memory and not stored on disk in a Sesame-specific format like the Native Java Store repository [54].

#### 4.1.3 RDFLogger

The RDFLogger component is used to transform an XLIFF document into RDF statements. This transformation from an XLIFF document to RDF statements is done through XSL

transformation (XSLT) which processes the XLIFF document to create Notation3 (N3) statements. N3 statements is another way of modelling RDF's XML syntax. N3 makes RDF's XML syntax more compact and human readable. The RDF statements are then merged into the triple store. The triple store repository URL and repository name that the RDF statements are merged into are stored in the config.properties file in the RDFLogger servlet. The entries in the config.properties file that are edited for this project are:

- DFLT SESAME URL - <http://killer7.s20.eatj.com/openrdf-sesame/>
- DFLT REP ID - NewCNGL

#### **4.1.4 Visualiser**

This Visualiser component was built by Sebastian Moliness as part of a CNGL funded project. This component is used to visualise data that has come from the triple store. The data from the triple store is matched to XLIFF phase names (in a JSON-encoded definition in file definitions.js), and the front-end then displays the RDF visualization. This component is really useful in the visualisation of data from the triple store as it helps to show a graph of the data from the triple store.

The Visualiser uses CURIEs (compact URIs) to display URIs in a compact form. For this purpose, rather than maintaining a set of namespace definitions (which would be difficult to maintain) it retrieves the current list of namespaces from Sesame Server. Adding or modifying namespaces is done through Sesame (typically via the Namespaces page in the Sesame Workbench web application).

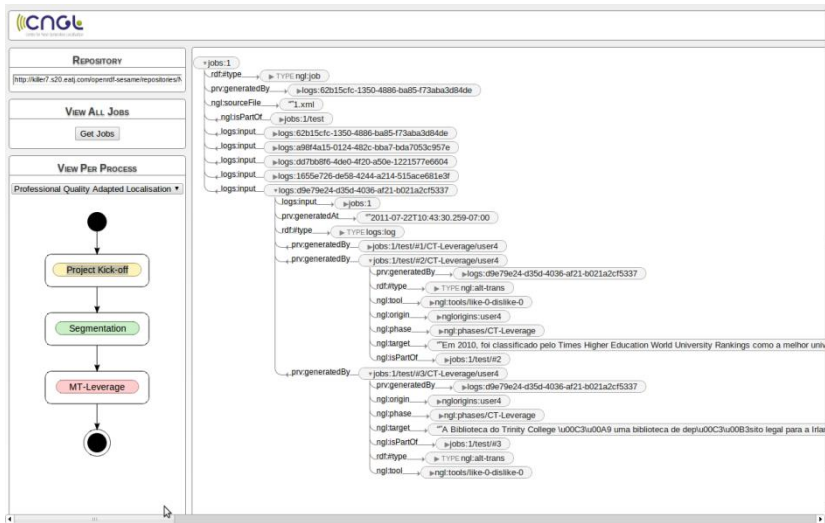


Figure 4.3: Visualiser

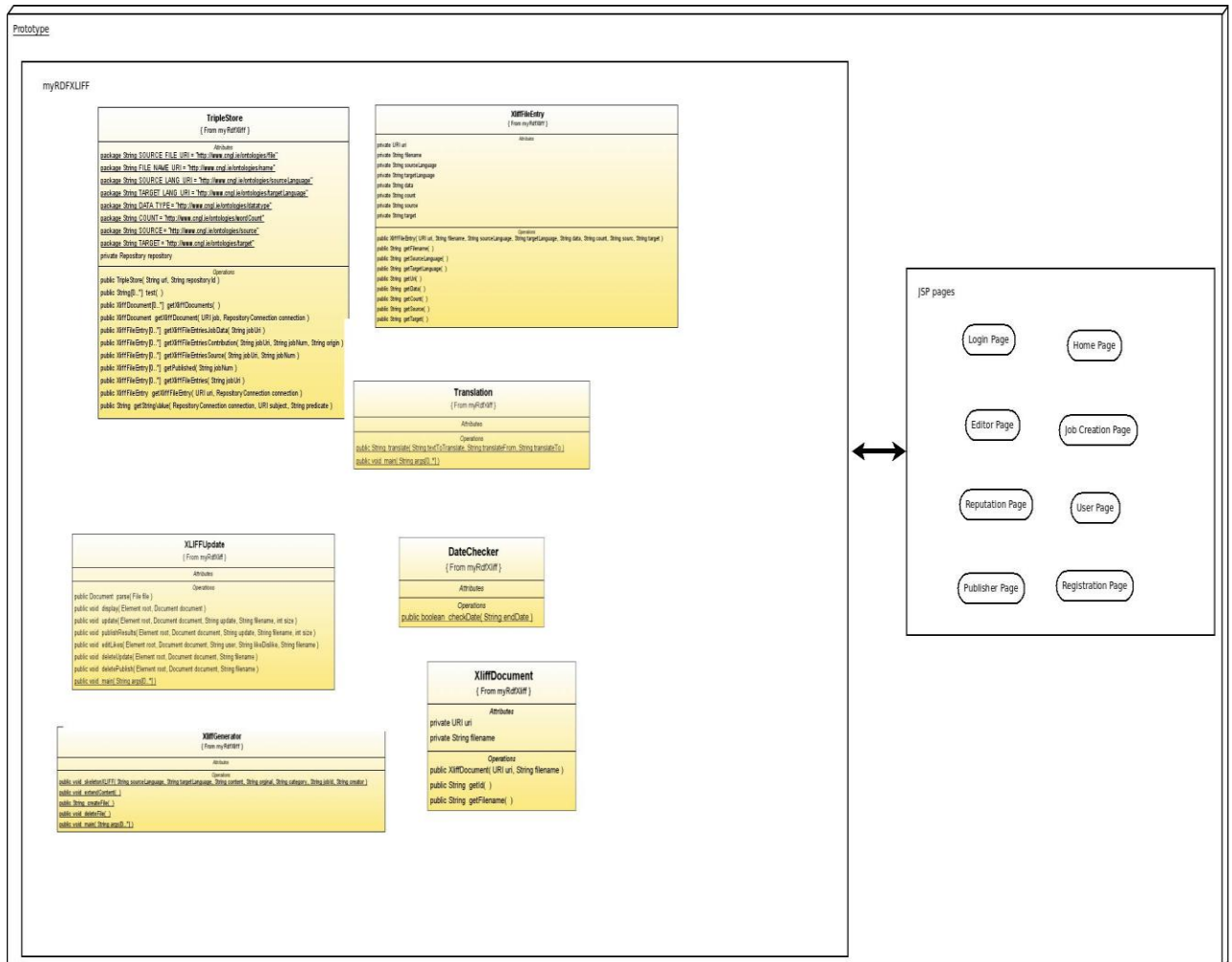


Figure 4.4: Prototype Architecture

#### **4.1.5 Prototype**

The Prototype component was the main part developed in this project. This component is the mobile crowdsourcing localisation application. The application will as discussed in the design section will offer users of the Prototype the opportunity to gain the knowledge of the crowd in relation to translation/localisation of content through the use of mobile devices as well as being usable on desktop computers. The application allows user to create user accounts, create localisation jobs, contribute to localisation jobs, comment on user content, rating user content and publish localisation jobs, these aspects will be discussed in more detail in section 4.2.

All the localisation content in the application is uploaded to the triple store on the Sesame server using the RDFLogger. Similar, to the RDFLogger the Prototype has a config.properties file which points towards the Sesame server and the repository where all the localisation content is held . The Prototype requires this configuration file because it needs to read data from the triple store. The prototype is required to read data from the triple store because it needs to display localisation data stored from the triple store to the users of the application. The entries in the config.properties file are configured like so:

- DFLT SESAME URL - <http://killer7.s20.eatj.com/openrdf-sesame/>
- DFLT REP ID - NewCNGL

The data is retrieved from the triple store using SPARQL queries. SPARQL is a query language for RDF data [71] and it allows for queries to consist of triple patterns, disjunctions and conjunctions patterns as well as optional patterns. The implementation of SPARQL used in prototype was done using Java libraries provided by the Sesame framework. The MySQL database holds all other data such as user data, job data, comment data and rating data in the Prototype (see section 3.3.6 for more details).

## **4.2 Details of Implementation**

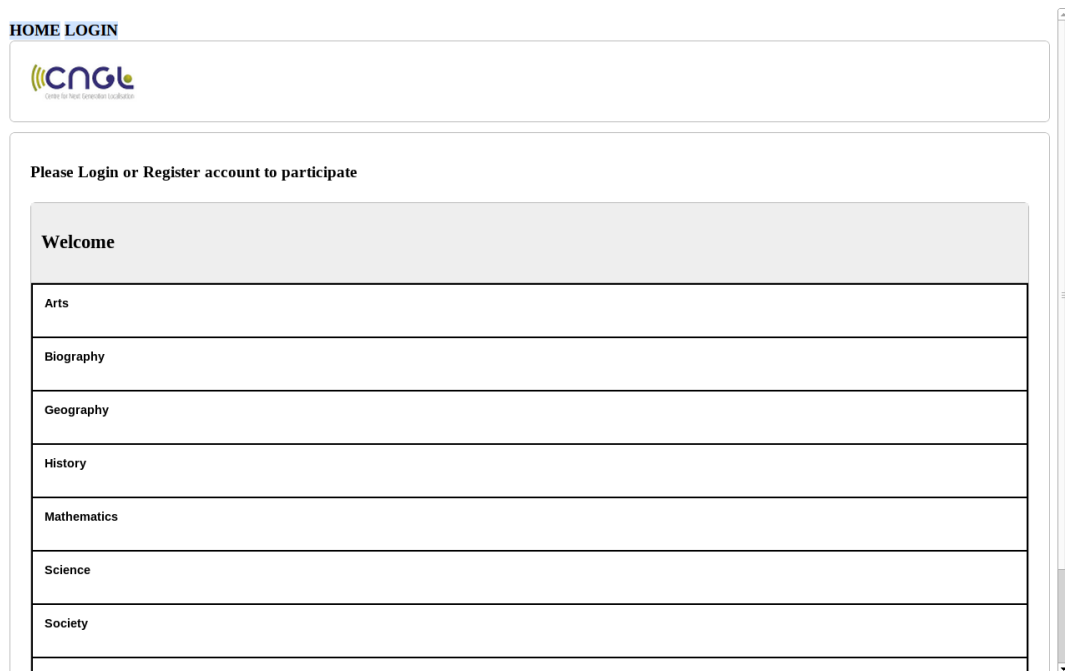
### **4.2.1 Home Page**

An image of the home page can be seen in Figure 4.5. The home page is the main page of the application and it shows the various categories in which jobs can exist in the prototype. The categories listed are Arts, Biography, Geography, History, Mathematics, Science, Society and Technology. The category names become clickable if the user is logged in. Otherwise, the



category names are not clickable if the user is not logged in to the application. If the user is not logged in then the home page recommends the user to log in or create a user account.

A bar at the top presents a link to the home page and a link to the login page which are visible if the user is not logged in. If the user is logged in the bar at the top will display a link to the home page, a link to the user page of the user who is logged in, a link to a page for users to create new jobs, a link to the reputation board which shows the register users in the application and the users reputation score and a link to log out of the application.



**Figure 4.5: Home page**

#### **4.2.2 Login Page**

The login page is used so that registered users of the application can login so that they can access their user account and can contribute translations, comments and ratings to the community. If a user log in is unsuccessful then the user is prompted that they have entered an incorrect user name or password. If a user log in is successful the user will be brought to their user page (section 4.2.4). If a user doesn't have a user account registered they cannot contribute translations, comments and ratings to the community. To get a user account all a user has to do is click the registration link on the login page and it will bring the user to a registration page (section 4.2.3).

**User Login**

**User Login**

User Name:	<input type="text"/>
Password:	<input type="password"/>
	<input type="submit" value="Submit"/>

[REGISTER NEW USER](#) This is a link to a page on this website.

**Figure 4.6: Login page**

### JSP session for login

Once a user logs in, it is important to associate information with the user such that data can be passed on from page to page in the prototype. This is done in JSP via “session”s which is an object that can be associated with a user. The session object is used to put data into and to retrieve data from much like a Hash table. Each user that logs into the prototype gets their own session with its own set of data specific to that user. This is very useful for storing and passing on of data from page to page for example storing the current user that is logged in.

```
session.setData();
```


```
session.getData();
```

### 4.2.3 Registration Page

The registration page is used to register new users to the application. The registration page is only accessible if the user of the application is not logged in. The link to the registration page can be found on the login page (section 4.2.2). The registration page is in a form format and asks the user that wants to register a few details so the account can be set up. The details asked for in

the registration form are the users first name, last name, email address, user name (each user name must be unique and not like any other registered user), and the password for the users account. If any of the details are left blank or have been entered incorrectly then the registration page with all the details filled is shown again and the user will have to edit the page depending on the incorrect details. The registration page may be shown again as well if the user name given by the user for registration has already been registered. This is because all user names must be unique and not already be registered in the application. All users that register as a user of the application have their user details stored in a MySQL database. The user details are stored using MySQL in the user table and the users interests are stored in the interest table.

[HOME LOGIN](#)



**User Registration**

First Name:	<input type="text"/>
Last Name:	<input type="text"/>
User Name:	<input type="text"/>
E-mail address:	<input type="text"/>
Interests:	<input type="checkbox"/> Biography <input type="checkbox"/> Geography <input type="checkbox"/> History <input type="checkbox"/> Mathematics <input type="checkbox"/> Science <input type="checkbox"/> Arts <input type="checkbox"/> Society <input type="checkbox"/> Technology
Password:	<input type="password"/>
Re-Enter Password:	<input type="password"/>
	<input type="submit" value="Submit"/>

**Figure 4.7: Registration page**

#### 4.2.4 User page

The user page is used to help organise the user jobs into different categories and to show the details about their user account such as their current reputation rating in the community. There are four different categories in which jobs can be organised into which are:

- **My jobs** - these are the jobs created by the user that is currently logged in which also haven't reached their completion date yet.

- **Public User Jobs** - are translation jobs which have been publicly given to all users of the application for translation. This means that all registered users can work on these jobs.
- **Private jobs** - are translation jobs which have been privately sent to a handful of user by the creator of the job. This means only users that have been sent the job can work on it.
- **Old jobs** - these are jobs that have reached their completion date and are now only visible to the creator of the job. The old jobs cannot have any more work done on them. The creator of the job can look at the results of the old jobs and see all translation contributions given by users along with the ratings of these contributions. The comments of other users about the job can be also seen along with the rating of the users comments. These old jobs are discussed in more detail in the publisher page section **pub**.

In each category, there is a table about the details of the jobs. The table that is displayed in each category displays data such as the jobs name, a job description, the user name of the person that created the job and the date that the job will end. The details about the jobs in the application are retrieved from the Jobs table in the MySQL database. The table is then evaluated and the required jobs go into the correct categories. For example distribution of jobs that are private that are in date go into the private job category, jobs that are public that are in date go into the public user jobs category, jobs created by the current user that is logged and that are in date in go into the my job category and jobs that are out of date create by the user that is currently logged in go into the old jobs category.

Users can select jobs to work on by simply clicking the link to the job which is located in the job name portion of the table. The users reputation score is located near the top of the user page and is labelled reputation score. An example of a user page can be seen in figure 4.8.



## Jobs

## User Account

User Rating : 3

## My Jobs

Job Name	Job Description	Username	Date Ended
KERBEROS PROTOCOL:BACKGROUND	Please translate this to French	DEMO	31.08.2011
MUSIC ARTICLE	I need it translated to German	DEMO	31.08.2011
TEST	This is a test	DEMO	31.08.2011

Public User Jobs

Figure 4.8: User page image

#### 4.2.5 Job Page

The job page is used to allow members of the community to create new translation jobs. To create a new job in the application the user must give details about the job that they want to create. The details that users have to enter in the job page to create a new job are:

- **Job name** - the name of the job and it doesn't have to be unique.
- **Job description** - a brief description of the job that has to be created.
- **End date** - this is end date or the date of completion marked for the job. After this date the job will only be available to the creator of the job. The end date is selected via a JavaScript calendar.
- **Translate from** - this is the language that the source content has to be translated from. The language to translate from is selected from a drop down menu which has the following languages available English, French, Italian, Spanish, German, Portuguese, Dutch and Romanian. As discussed in the project scope there are only certain languages available for the prototype.

- **Translate to** - this is the language which the source content has to be translated to. The language to translate to is selected from a drop down menu which has the following languages to choose from English, French, Italian, Spanish, German, Portuguese, Dutch and Romanian.
- **Category** - the category of the job is selected by the user and it should be based on the source content that has to be translated. The category of a job is selected via a drop down menu which has the following options which are Arts, Biography, Geography, History, Mathematics, Science, Society and Technology.
- **Select Job Distribution** - the job distribution can either be public or private. If the job is public then the translation job being created will be visible to all registered users of the application. If the job is private the translation job will only be visible to the creator of the job and the users that were sent the job (**Send to**).
- **Send to** - this part of the form is used so that the creator can send the job to particular users in the application. The creator of the job can send the job to as many users as they want. If the job is private then it is essential that the creator of the job fill in the send to list or they will be the only person to receive the job.
- **Source content to be translated** - this is the content that has to be translated for the job.

Once the user has filled in the details of the job they can submit it and there details are written to a submission page where the creator of the job can review the details of the job that they have created. Once they click submit on this page there data is submitted and the job is now in the application. The data about the newly made job is divided up and stored in two places which are the MySQL table and the triple store. The MySQL job table stores the jobs name, job description, the end date, the user that created the job, the category of the job, the distribution of the job (public or private) also the table assigns the job a unique identification number. The dist MySQL table stores the user names of the users who have been sent the job if there are any.

The source content that needs to be translated will be kept in an XLIFF document. Every job that is created by a user has an accompanying XLIFF document which is created on the Apache Tomcat server. The XLIFF file format is an extension of XML which was created to standardize localization. file content. The files are named based on an incremental process whereby the file is named based on the number of other jobs that have been created in the application. For example say there was 100 jobs created in the application and a user creates a new job, the name of the file for the new job is “101.xlf”. The number of jobs in the application is calculated by the

number of jobs in the Jobs table of the MySQL database. This number is gotten by using the JDBC Java library to get the number by the use of the following code:

```
stmt.executeQuery("select count(*) from jobs");
```

### **XLIFF structure**

The XliffGenerator Java class is used to generate the XLIFF document based on the data entered by the user when they are creating a job. The *skeletonXLIFF* method in the XliffGenerator creates the XLIFF document for the new job. The data about the job written to the XLIFF file is:

- the source content that need to be translated,
- the language to be translated to,
- the language to be translated from,
- the creator of the job,
- the jobs identification number,
- the category of the job,
- the file name,
- the amount of characters in the source content,
- the amount of words in the source content,
- the amount of sentence in the source content and
- the source content is split up into sentences and are added into the XLIFF document in separate trans-unit section and are labelled according to the sentence number for example the first sentence will be 1 and the second sentence will be 2 etc .

The source content is segmented into individual sentences and each of the sentences is put into a different trans-unit dependent on the order of the sentence. For example the first sentence will go into trans-unit “#1”, the second sentence will go into trans-unit “#2” and this continues until all sentences from the source content are placed in a trans-unit. Then within the trans-units there are multiple alt-trans which have an origin element within them. The origin element for in the alt-trans tag can have three possible types which are:

- sentence - the source sentences that needs to be translated in the job.
- user(number) - this sentence type is the content translation given by users and it is incrementally numbered such that if the last user translation was user1 the next user who submits a translation will be user2.

- published - this is the translation that has been chosen as the best sentence to be published by the users.

The type to use in the situation of source content is sentence which indicates that the sentence is a piece of source content of the job and it has to be translated by users of the application. A sample of a source content sentence in the XLIFF document can be seen in Figure 4.9 below:

```
<trans-unit id="#1">
  <alt-trans tool-id="" origin="sentence" phase-name="">
    <target>This program features Lech Kowalski's 2003 documentary feature about the life
and times of Ramones bassist</target>
  </alt-trans>
</trans-unit>
```

**Figure 4.9: Sentence type in XLIFF**

Once the data has been written to the XLIFF document and it is located on the server then the RDFLogger servlet transforms the XLIFF document into RDF statements. The transformation from XLIFF to RDF statements is done using XSLT which processes the XLIFF file to produce N3 statements. Once these RDF statements have been produced they are then merged into the Sesame triple store.

**Figure 4.10: Job creation page**



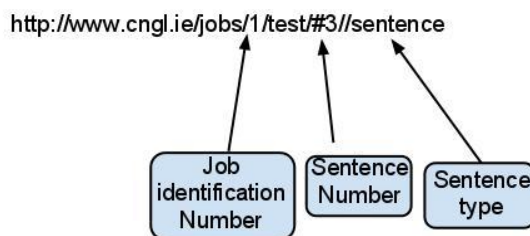
#### 4.2.6 Editor Page

This page is where users in the community can see a particular translation job depending on the job link that the user clicked on. The job identification number is unique to each job and is required in the retrieval and insertion of data into jobs in the prototype. The jobs identification number is encoded as a query string in the URL for all job links in the prototype. The question mark in the URL is used as a separator and it is used to indicate the start of the query string. An example of the URL with a query string to a particular job is **.../prototype/editor.jsp?test-24**. The end of the URL contains the job number which is 24.

##### 4.2.6.1 Resource Identification

The editor page of the application requires localisation data relating to a particular job to be extracted from the triple store. This data is read from the triple store through the use of SPARQL queries. These SPARQL queries are created using the Java Sesame libraries. Every job in the triple store is uniquely identified by its job identification number, this makes it possible to determine when a statement in the triple store is related to a particular job.

In the RDFLogger component the XSLT transforms the XLIFF document to RDF statements. The XSLT constructs URIs based on the unique job identification number to which names are appended in a manner that matches the structure of the XLIFF document . An example of the URIs constructed in the triple store can be seen in the Figure 4.11 below:



**Figure 4.11: URI sample**

The URI in Figure 4.11 uses the job identification number to identify the unique job, the sentence number identifies the sentence number in the store and sentence type refers to what type of sentence is being referred to in the triple store and in the case of Figure 4.11 it is a source

sentence. The Figure 4.11 for example refers to the third sentence of the source content from job 1.

#### 4.2.6.2 SPARQL SELECTS

The class that queries the triple store by way of SPARQL queries in the implementation of the prototype is called TripleStore. The type of SPARQL queries that are used to get data from the triple store are called SELECT queries. There are three main SELECT queries used to retrieve data in the editor.jsp and publisher.jsp page which are as follows:

##### Source content

The query that gets the source content that needs to be translated from the triple store is as follows:

```
SELECT ?sub WHERE
{
?sub <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.cngl.ie/ontologies/alt-trans> .
?sub                                <http://www.cngl.ie/ontologies/isPartOf>
<http://www.cngl.ie/jobs/"jobNum"/test/"sentenceNo">.
?sub <http://www.cngl.ie/ontologies/origin> <http://www.cngl.ie/ontologies/origins/sentence>.
}
```

The SELECT query looks at the data that is **alt-trans** type in the triple store. The **jobNum** or job identification number of an existing job is selected and it will get the appropriate job. The sentence number **sentenceNo** starts at one and goes incrementally through each of the trans-units until it gets all the source sentences for the job and then these source sentences are displayed for the users. The function in the TripleStore class that gets this data is called **getXliffFileEntriesSource**.

##### User contribution content

The next content that needs to be retrieved for the triple store is the translation content that has been contributed by the crowd. The query that gets the user contribution content from the triple store is as follows:

```

SELECT ?sub WHERE {
?sub <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.cngl.ie/ontologies/alt-
trans> .
?sub                                     <http://www.cngl.ie/ontologies/isPartOf>
<http://www.cngl.ie/jobs/"+jobNum+"/test/"+jobUri+">.
?sub<http://www.cngl.ie/ontologies/origin><http://www.cngl.ie/ontologies/origins/user"+userC
ounter+">.

```

The selecting of user contributed content is similar to getting the source content. The query gets the data from the **alt-trans** type of the triple store where the **jobNum** or job identification number of an existing job is selected and it will get the appropriate job. The sentence number **sentenceNo** starts at one and goes incrementally through each of the trans-units until it gets all the content translation given by a user. Then the user **userCounter** is incremented by one to get the next users data and the **sentenceNo** is set back to 1. This loops until all user contributions from user 1 to n is gotten from the triple store. This data is then displayed in editor.jsp and publisher.jsp to users of the application. The function in the TripleStore class that gets this data is called **getXliffFileEntriesContribution**.

Every user translation contribution has a user name of the person that gave the contribution and the rating score associated with it. This data is retrieved from the Translation MySQL table based on the job number and the reference number to the contribution. This rating and username information for the contribution are displayed beside the associated user contribution as seen in Figure 4.13.

### **Job Data content**

The job data content that needs to be retrieved from triple store is data such as the language to translate data to and the language to translate data from. The following query will get all job data about a particular job apart from the language data such as translation and source content from the triple store:

```

SELECT ?sub WHERE {
?sub <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.cngl.ie/ontologies/file>.
?sub <http://www.cngl.ie/ontologies/isPartOf> <http://www.cngl.ie/jobs/"+jobNum+">.
}

```

This query gets the data from the **file** type in the triple store and gets the job data based on **jobNum** or job identification number which selects the appropriate job. The function in the TripleStore class that gets this data is called **getXliffFileEntriesJobData**. The translation languages are extracted from this job data and the language to translate to and from is displayed to the user at the top and near the bottom of the page Figure 4.13.

#### 4.2.6.2 Contributions

##### Job Contribution

There are two ways in which users can contribute to the crowd which is through regular translation of the entire content by themselves or they can contribute with the aid of machine translation with post-editing done by the user. The machine translation as discussed in the design section is done through a Java library which uses Microsoft translation called Microsoft-translator-java-api. The languages available for machine translation are English, Spanish, German, Italian, Romanian, Portuguese and Dutch which are within the project scope discussed in the design section. The Microsoft-translator-java-api is added as a jar file onto the project. The Translation class is where the translation take part through the translation function (**translate(String textToTranslate, String translateFrom, String translateTo)**). The translation function takes in the text that has to be translated, the language that has to be translated from and the language to translate to.

When a user submits a translation contribution by either method the XLIFF document is updated to include the new translation content from the user. The user contributed translation content is divided up into sentence and it is written to the XLIFF document into the section with its related source sentence. This means each source translation content sentence is represented in individual trans-unit along with the user contributed translations for that source sentence as seen in Figure 4.12. The origin element in the alt-trans tag is labelled incrementally so if the last person to contribute was user1 the next user to contribute will be user2 as illustrated in Figure 4.12.

```
<trans-unit id="#1">
  <alt-trans tool-id="" origin="sentence" phase-name="">
    <target>This program features Lech Kowalski's 2003 documentary feature about the life
    and times of Ramones bassist</target>
```

```

</alt-trans>
<alt-trans tool-id="" origin="user1" phase-name="">
<target>Este programa cuenta con 2003 Lech Kowalski documental sobre la vida y obra del
bajista Ramones </target>
</alt-trans>

<alt-trans tool-id="" origin="user2" phase-name="">
<target>Este programa cuenta con 2003 Lech Kowalski documental sobre la vida y obra
del bajista Ramones </target>
</alt-trans>
</trans-unit>

```

**Figure 4.12: User Contribution being added to the XLIFF document**

Once the updates have been written to the XLIFF document then the RDFLogger servlet can transform the XLIFF document into RDF statements. These RDF statements are then merged with the triple store so that the triple store now includes data from the newly updated XLIFF document.

The Translation table in the MySQL database is updated as well with an entry when a translation contribution takes place. This update to the Translation table consists of the user name of the person that contributed the translation, the job identification number of the job and the “like” and “dislike” rating is set to zero.

### **Commenting and Rating Contributions**

The users of the application can also contribute by commenting on the translation job and by rating user contributed translations or comments. The users can add on comments to the job simply by clicking “Add comment” and users can rate content by simply clicking the like or dislike button beside the comment or user translation.

HOME DEMO CREATE NEW JOB REPUTATION BOARD LOG OUT

**Source Translation Content**

1001 Albums You Must Hear Before You Die is a musical reference book edited by Robert Dimery, first published in 2005. The most recent edition consists of a list of albums released between 1955 and 2009, part of a series from Quintessence Editions Ltd. The book is arranged chronologically, starting with Frank Sinatra's In the Wee Small Hours and concluding with It's Blitz! by Yeah Yeah Yeahs.

User Translated Content			
Target	User	Like	Dislike
1001 Alben Sie hören, bevor Sie Must ist ein musikalisches Nachschlagewerk von Robert Dimery ediert, zum ersten Mal 2005 veröffentlicht. Die jüngste Auflage besteht aus einer Liste der Alben zwischen 1955 und 2009 veröffentlicht, gehört zu einer Reihe von Quintessence Editions Ltd Das Buch chronologisch angeordnet, beginnend mit Frank Sinatra's In the Wee Small Hours und den Abschluss mit seinem Blitz! von Yeah Yeah Yeahs.	DEMO	LIKE -0	DISLIKE -0

**Translation toolbar**  
English--->German

- 1001 Albums You Must Hear Before You Die is a musical reference book edited by Robert Dimery, first published in 2005
- The most recent edition consists of a list of albums released between 1955 and 2009, part of a series from Quintessence Editions Ltd
- The book is arranged chronologically, starting with Frank Sinatra's In the Wee Small Hours and concluding with It's Blitz! by Yeah Yeah Yeahs

**REGULAR TRANSLATION TRANSLATION COMMENTS**

Text to be translated: 1001 Albums You Must Hear Before You Die is a musical reference book edited by

From: English

To: English

Translated Text:

[German](#) [Spanish](#) [Italian](#) [French](#)

Figure 4.13: Editor page

### Keyboard support

As discussed in the design section keyboard support for some languages is lacking in smart phones such as remote local dialects. So for this dissertation keyboard support will be prototyped for languages that are within the scope of this project which are Spanish, German, Italian, Romanian, Portuguese and Dutch. The keyboards for the different languages will be available by clicking on the language links to the keyboards which can be seen in Figure 4.13. The keyboard was made using a combination of JavaScript and HTML. The background work of getting the letters that are entered and the use of tab and space buttons is done using JavaScript. The keys on the keyboard are displayed using HTML.

### 4.2.7 Publisher Page

The publisher page is used to display jobs that have reached their end date and are now only visible to the creator of the job. The creator of the job cannot edit any of the users contribution or add in extra contributions to the job but the creator can do the following:

- analyse the source content that they wanted translated,
- analyse the translation contributions from the community along with users names of the people that contributed along with the ratings (likes and dislikes scores) that the translations received from the community,
- analyse other users comments on the jobs along with viewing the ratings which the comments have received and
- the creator of the job can select and publish the best user translation which will increase that users reputation score in the community.

The publisher page has similar layout to the editor page. The way the job data, source translation data and the user contributed translation data is selected from the triple store and MySQL is done same way as it was for the editor page and a description of this selecting of data can be found in section 4.2.6.2. There is a publish option beside each of the user translation contributions. If the creator of the job clicks on publish then this user translation has been selected as the best and then goes to the publishing submission screen. The user on this submission screen has two options which is to publish or not publish. If the user selects not to publish then they are brought back to the publisher page Figure 4.14.

HOME DEMO CREATE NEW JOB REPUTATION BOARD LOG OUT

The screenshot shows a web interface for a translation project. At the top, there are navigation links: HOME, DEMO, CREATE NEW JOB, REPUTATION BOARD, and LOG OUT. Below this is the CNGL logo. The main section is titled "Translation units" and shows "English-->Italian". Underneath, there is a "Source Translation Content" box with a yellow background containing text about Leonardo's early life. Below that is a "User Translated Content" table with two rows of translations. Each row includes a "Target" column, a "User" column, "Like" and "Dislike" columns with counts, and a "PUBLISH" button. The first row is by user "DEMO" with 2 likes and 0 dislikes. The second row is by user "FINNLE" with 1 like and 1 dislike. At the bottom, there is a "Published Translation" section, also showing "English-->Italian".

User Translated Content				
Target	User	Like	Dislike	
Poco si sa di primi anni di vita di Leonardo.Ha trascorso i suoi primi cinque anni nella frazione di Anchiano a casa di sua madre, poi dal 1457 viveva nella casa di suo padre, i nonni e lo zio, Francesco, nella piccola città di Vinci.Suo padre aveva sposato a sedici anni, ragazza di nome Albiera, che amava Leonardo, ma morì giovane.[10] Quando Leonardo sedici anni suo padre si risposò, a venti anni, Francesca Lanfredini.Non è stato fino al suo matrimonio terza e la quarta che Ser Piero prodotto eredi legittimi.	DEMO	-2	-0	PUBLISH
Poco si sa circa i primi anni di vita di Leonardo.Ha trascorso i suoi primi cinque anni nella frazione di Anchiano nella casa di sua madre, poi dal 1457 visse nella casa del padre e lo zio, Francesco, nella piccola città di Vinci, nonni.Suo padre aveva sposato una ragazza di sedici anni nominata Albiera, che amava Leonardo ma morì giovane.Quando Leonardo aveva sedici anni suo padre si sposò di nuovo, a venti anni Francesca Lanfredini.Non fu fino ai suoi terzi e quarto matrimoni che Ser Piero prodotto eredi legittimi.	FINNLE	-1	-1	PUBLISH

Figure 4.14: Publisher page

Otherwise, if the user selects publish the XLIFF document is updated by the adding of another trans-unit to the XLIFF document as seen in figure 4.15. The id of the trans-unit in the other parts of the document are labelled by numbers which represents the sentence number. The id for the new trans-unit for the published translation is published and the trans-unit section gets an alt-trans section with an element origin which represents the sentence type to be published. The data contained in between the alt-trans section is the entire translation that has been decided to be the best and therefore be published. This new trans-unit section which contains the published data is then merged into the triple store using the RDFLogger. The user MySQL table for the job is updated such that the published column in the job now is set to true and the user who contributed the translation has their reputation score increased by five in the user MySQL table.

```
<trans-unit id="published">
  <alt-trans tool-id="" origin="published" phase-name="CT-Leverage">
    <target>Este programa cuenta con 2003 Lech Kowalski documental sobre la vida y obra
del bajista Ramones y All-Star se queman, Dee Dee Ramone (1952-2002).La vida de Dee Dee es
un estudio fascinante personaje de una leyenda del punk rock que nunca crecí3.Escuchar a la
cuenta de Dee Dee con el director Lech Kowalski, para discutir Johnny Thunders de la
película &quot;Born to Lose&quot;.</target>
  </alt-trans>
</trans-unit>
```

**Figure 4.15: Publishing Job XLIFF**

The bottom of the published page now contains the published translation that was selected by the creator of the job. The SPARQL query to get this data is as follows:

```
SELECT ?sub WHERE {
?sub <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.cngl.ie/ontologies/alt-
trans> .
?sub<http://www.cngl.ie/ontologies/isPartOf><http://www.cngl.ie/jobs/"+jobNum+"/test/publis
hed>.
?sub<http://www.cngl.ie/ontologies/origin><http://www.cngl.ie/ontologies/origins/published>.
```




The query gets the data from the **alt-trans** type of the triple store where the **jobNum** or job identification number of an existing job is selected and it will get the appropriate job where the translation unit identification is **published**. The sentence type is then retrieved which is of type **published** and then the selected query data is displayed to the user at the bottom of the publisher page. If there is no published data in the triple store then the bottom area where the translation data is displayed will be blank.

#### 4.2.8 Reputation board

The reputation board is used to show the users that are currently registered members of the community. The reputation table shows the user names, the reputation scores of the users and the users interest in the community. This page acts as a sort of community page because it allows members of the community to look at other users interests along with all the users reputation scores. This allows additional sociability and visibility elements in the application leading to a more well informed community because community members can see which users have the best reputation. Also users can click on other user names to see other user pages and see what jobs other users have created in the prototype

HOME DEMO CREATE NEW JOB REPUTATION BOARD LOG OUT



### Reputation Board

Username	Rating	Interests
<a href="#">FINNLE</a>	5	Technology Society Arts Biography
<a href="#">KILLER7</a>	1	Technology Society History Biography
<a href="#">ASDA</a>	4	Technology Mathematics Geography
<a href="#">ASDAMAN</a>	2	Technology Society Biography
<a href="#">EMER</a>	3	Technology Society
<a href="#">ADAM</a>	0	Technology Science Mathematics
<a href="#">GUXU</a>	0	Arts Science Mathematics
<a href="#">DARWIN14</a>	0	Technology Arts Science Mathematics History
<a href="#">LEE FINN</a>	0	Technology Science Biography
<a href="#">DLEWIS</a>	0	Technology Science
<a href="#">DAVID.FILIP@ULIE</a>	0	Technology Society Arts Science Mathematics History Geography Biography
<a href="#">RAJATGUPTA</a>	0	Science Mathematics
<a href="#">SCHORSI</a>	1	Technology Society Science
<a href="#">TERRANIGMA0</a>	12	Technology Society Arts Science

Figure 4.16: Reputation board

# Chapter 5

## Evaluation

This section will evaluate the prototype that was discussed in chapter 4. This chapter discusses the evaluation process that took place in this dissertation from the drawing up of evaluation requirements to the pilot studies results followed by the main studies results.

### 5.1 Evaluation Approach

#### 5.1.1 Evaluation Requirements

At the early design phase of the application a set of evaluation requirements were devised. These evaluation requirements informed the approach for prioritizing features in the design phase and implementation phase of the prototype. This prioritizing of requirements was very important because it helped in focusing the design such that the development of wasteful superficial elements which had no bearing or contribution to the evaluation of the prototype were not developed. The list of final evaluation requirements are as follows:

- The evaluation of the usability of the prototype
- The evaluation of the allowance for community deliberation and self management in the prototype
- The evaluation of the sociability elements in the prototype
- The evaluation of the reputation and visibility of users in the prototype
- The evaluation of the suitability of this type of application on a mobile device

#### 5.1.2 Evaluation Scoping

The evaluation of the prototype will take place through limited user usage studies. Therefore some of the above evaluation requirements will be hard to evaluate correctly due to the large amount of data that will be required to accurately do an evaluation. Some examples of such evaluation requirements which will need a sufficient amount of data are:

- evaluating the usability of users adding translations to other users jobs,
- evaluating usability requirements such as the testing of publishing of jobs which will require out of date jobs to evaluate,

- evaluating of social and usability requirements such as rating of other users content and commenting on jobs in the application.

Therefore some fake data will be loaded into the application prior to testing of the application so that effective evaluation of these aspects of the application can take place.

The analysing of two way interactions in the application is deemed out of the scope of this project and will therefore be limited to analysing its usability in the application and not analysing the social aspect in the prototype. This is due to the amount of volunteers (twelve) and the amount of time that these volunteers will spend with the prototype (about an hour) during the testing will not be sufficient enough to build up the data required to do such an analysis.

### **5.1.3 Fake data**

Due to the amount of data that was needed as discussed in section 5.1.2 some fake data was set up in the application prior to user testing. The fake data that was set up in the application was three pretend users that created three fake jobs along with additional content for these jobs such as translations, comments and some ratings. On top of this, fake out of date jobs were created for the volunteer testers. This was done so that the volunteer testers could test out the usability of publishing out of date jobs in the application. For these reasons and so that the volunteers could have access to the out of date jobs the volunteers were sent a user name which they were able to use to create a user account under.

### **5.1.4 Evaluation process**

The evaluation took place on twelve participants who were bi-lingual and were also owners of a smart phone device or a tablet. The participant volunteers were a mixture of 80% males and 20% females and in the following text participant identities will not be used. The participants will be referred to as subject 1 to subject 10 and the numbering of the participants will not match or represent the order in which they took part in the evaluation. The evaluation was designed so that it would take less than an hour to complete. In accordance with Trinity College Dublin's research procedures, ethical clearance was received from the ethics committee before testing of the prototype took place on the participants. Every participant prior to testing had to sign off on

a consent form which informed the users about the project, the task that had to complete as part of the evaluation as well as highlighting the anonymous nature of the evaluation.

Once the participants had agreed to the conditions on the consent form the participants then could take part in the evaluation. The participants were given a worksheet which contained the details of the exercises that they had to complete in order to take part in the evaluation of the prototype . The worksheet can be seen in Appendix D. The worksheet detailed the following tasks for participants to complete:

- Create a user account
- Create a translation job
- Contribute translation content to a translation job
- Rate other users contributions
- Comment on a translation job
- Publish a translation job

In completing these exercises on the worksheet the participants using a mobile device were able to form an opinion on the prototypes various components such as the collaborative, social and self-management components of the prototype. Therefore the participant will be able to assess the usability and suitability of the prototype for a mobile device.

Once the participants had finished the worksheet they were asked to answer 21 questions on a questionnaire which was hosted via Google Docs. The questionnaire (Appendix B) was split into four parts which are:

1. System Usability Scale (SUS)
2. Sociability Support
3. Application Analysis
4. Participant Profile

#### **5.1.4.1. System Usability Scale (SUS)**

The System Usability Scale section is a set of ten standard questions which help in evaluating the usability of the prototype. Good usability in a web application is important for user uptake of the application especially so in the context of a collaborative online application. The ten SUS

questions are answered via a five point rating scale system (likert scale), which goes from “*Strongly disagree*” (1) to “*Strongly agree*” (5). The ratings given in the questions help in calculating a SUS score which ranges from 0 to 100. The higher the SUS score the better the usability of the application.

The SUS score is calculated by summing the score of the contributions for each question. Every questions contribution ranges from 0 to 5. Questions 1,3,5,7 and 9 have a score contribution scale from the -1 position. Questions 2,3,6,8, and 10 have a contribution scale from the - 5 position [25]. Once you have the sum of the scores multiply the scores by 2.5 to get the overall System Usability score.

#### **5.1.4.2. Collaboration Analysis**

The second section the collaboration analysis is composed of five questions. The aim of the questions in the collaboration analysis is to see how effective the design decisions that were made in the prototype. This was done by gathering the participants opinions on:

- How well the prototype supports the emergence of contributor reputation,
- How well the transparency of participants contributions are handled,
- How well the expression of sociability is handled,
- How well the prototype supports making decisions on contributions in the application
- How well the prototype supports different types of distribution.

The five collaboration analysis questions were answered in a similar fashion to SUS question by the use of a five point rating system which also went from “*Strongly disagree*” (1) to “*Strongly agree*” (5).

#### **5.1.4.3. Application Analysis**

The Application analysis section is made up of three questions which ask the participants about their impressions of the prototype, is the prototype suitable for a mobile device and did they complete the all the tasks given to them in the worksheet (Appendix D). These questions will help to see participants opinion of the prototype whilst also seeing what their experience was of using the application by seeing how successfully they think they completed the worksheet. Questions 16 and 18 are answered via a five point rating scale which goes from *Strongly Negative* (1) to *Strongly Positive* (5). Question 17 is answered in the same fashion but the rating scale goes from *Strongly Negative* (1) to *Strongly Positive* (5).

#### **5.2.4.4. Participant Profile**

The participants profile is the last section of questions and it contains a total of 4 questions which are used to ascertain the participants experience in translation as well as seeing how often they contribute to online crowdsourcing websites. The level of translation experience helps to see how participants who are experienced in translation thought about the usability of the application and will be useful in analysing their evaluation results with this in mind. Participants level of contribution to crowdsourcing websites shows a participants familiarity to collaboration concepts which will be useful in analysing the participants evaluations. The last two questions ask two basic questions which are what mobile web browser did the participant use to do the evaluation followed by asking them for any feedback they may have about the prototype. Question 19 and 20 are answered again using a rating scale which go from “*Never*” (1) to the midpoint “*Occasionally*” (3) to “*Regularly*” (5). The last two questions 21 and 22 are just text boxes into which users can answer.

#### **5.2 Pilot Study**

A pilot study was carried prior to the main study. The pilot study is an experiment on a small group of participants to test out the logistics and to help gather data about the prototype prior to having the main study, in order to improve the quality and efficiency of the prototype [41].The pilot study took place one week ahead of the main study with two participants. This was done in order to allow time to make the required changes in the application which were required based on the results of the pilot study. The pilot study unlike the main study happened under supervised conditions and each of the pilot study participants were scheduled for one hour slots. The participants in the pilot study helped to highlight a few issues in the prototype which left unchecked could have possibly spoiled the evaluation results . The issues highlighted in the pilot study meant that modifications had to be made to the prototype. These modifications are discussed in detail in section 5.3.

#### **5.2.1 Pilot Study Findings**

The pilot study consisted of doing a trial run of the main study on two participants. This basically consisted of doing the exercises on the worksheet as seen in Appendix D followed by a questionnaire as seen in Appendix B. Both participants of the pilot study had a good overall impression of the prototype based on their questionnaire results. This section will present these questionnaire results given by both participants in the pilot study.

### 5.2.1.1. Pilot Study System Usability Scale

The questions in the SUS section are not analyzed individually but are analysed together to get a SUS score which goes from 0-100. This section will therefore focus on analysing the SUS score and the resulting pattern in the graph of the averaged results. The SUS score was calculated as discussed in section 5.1.4.1.

The questions in the SUS questionnaire alternate between positive (*e.g. I felt very confident using the system*) (question 1,3,5,7 and 9) and negative (*e.g. I found the system very cumbersome to use*)(question 2,4,6,8 and 10) . The average result from these questions are plotted onto a star graph which can be seen in figure 5.1. The figure 5.1 shows an even star shape which is a positive sign and shows that the two participants had an overall positive usability experience. Any negative response or inconsistency that occur in the SUS ratings would be clearly visible on the star graph by producing a misshapen star. The overall SUS score was also very positive at 72.50%. Both the graph and the SUS score indicate that both participants had a positive user experience whilst using the prototype which means that the prototype has overall good usability.

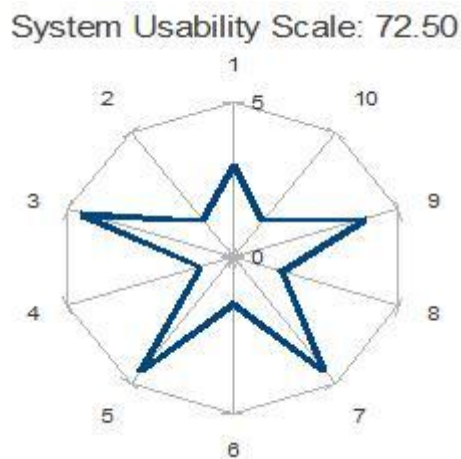
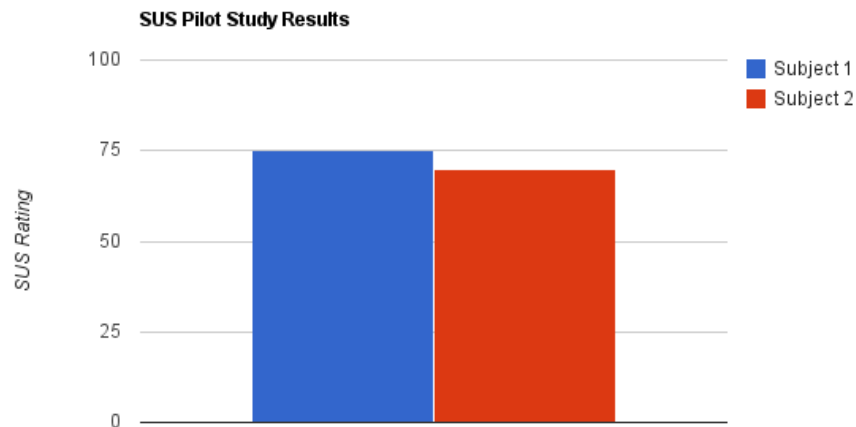


Figure 5.1: Pilot Studies System Usability Scale Average



**Figure 5.2: Pilot Studies SUS Results**

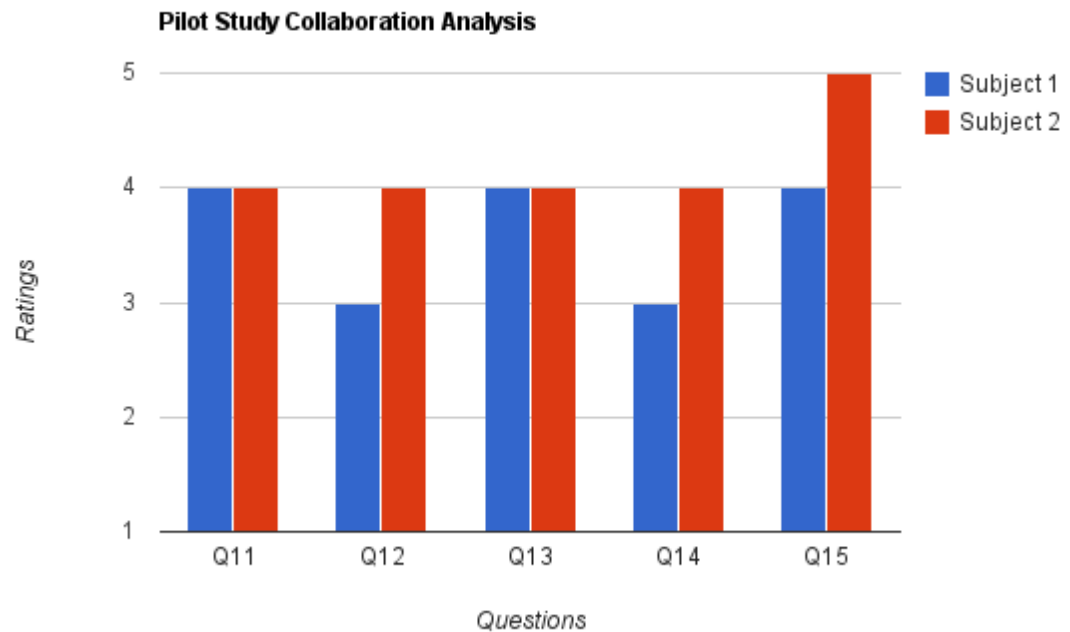
#### **5.2.1.2. Pilot Study Collaboration Analysis**

The second part of the questionnaire focuses on questioning how effective the collaboration components are in the prototype. The participants gave similar high marks for the collaboration features with none of the rating scores dipping below the rating of a 3 as seen in figure 5.3. The highest rating given was 5 for question 15 by Subject 2 which asks about the prototypes support for distribution of translation jobs in the prototype and Subject 1 gave it a high rating as well with 4. Question 11 and 13 were rated 4 by both participants in the pilot study. This high rating for question 11 shows that the participants thought the emergence of user reputation was well handled by the use of a reputation score. The equal high ratings in Question 13 shows that both participants thought sociability aspects of the prototype were well incorporated.

Question 12 and 14 were scored the same as each other with Subject 1 giving a rating of 3 for both and Subject 2 giving a rating of 4 for both questions as well. The high ratings in question 12 shows that there is a high confidence in the transparency of user contributions in the prototype. Whilst the high rating in question 14 shows that the participants opinion on user decision mechanisms is supported well in the prototype.



The average rating score for the collaboration features section in the questionnaire was 3.9 out of 5. This high average rating given by the participants shows a high satisfaction rating for the collaboration components in the prototype.



**Figure 5.3: Pilot Study Collaboration Analysis**

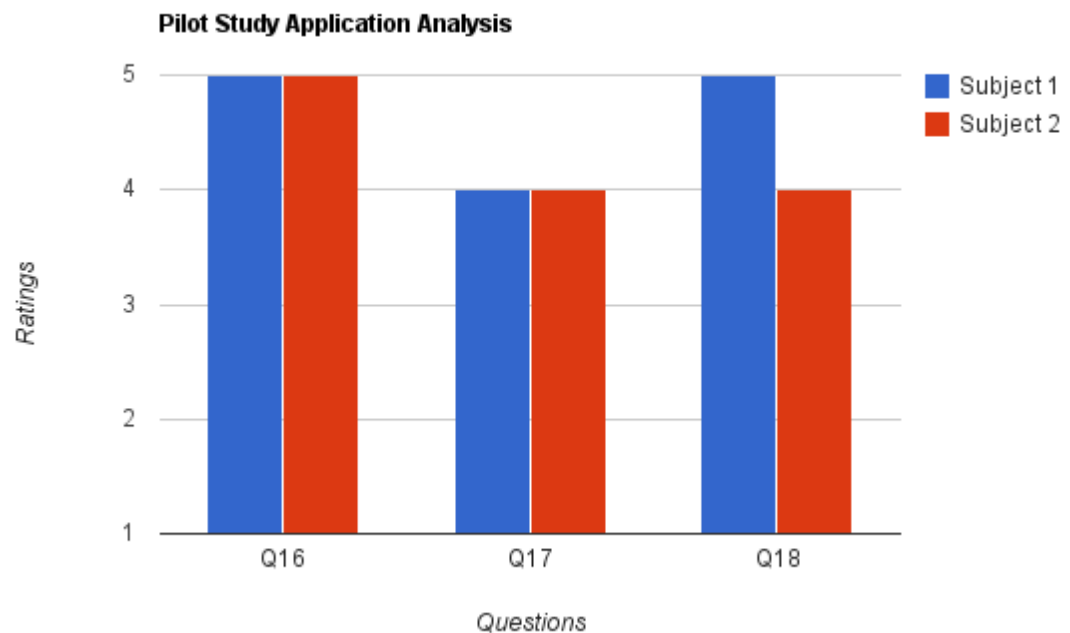
### 5.2.1.3. Pilot Study Application Analysis

The questions in the Application analysis were analysed separately and the questions are all positively phrased meaning the higher the answer rating the more positive the response. Both participants felt that they completed the task given to them successfully in the pilot study worksheet (Appendix D) as both participants gave strongly positive response of 5 to question 16. This means that the participants felt that they managed to complete all the exercises on the worksheet meaning that they can accurately give an evaluation for the pilot study.

The participants overall experience of using the application was strongly positive with both participants giving a positive rating of 4 to question 17. It was noted that some bugs in the application hindered some of the participants user experience. These bugs are discussed in more detail in section 5.2.2. The last question asked to the users in this section was question 18 which asked was the application suitable for a mobile device. This question scored very highly with the participants scoring a 5 and a 4 rating for this question. This means that the participants of

the pilot study think that this prototype of a crowdsourcing translation application for a mobile device is very suitable and worthwhile application to create.

The results for the Application analysis in the pilot study shows that both participants thought the prototype suits the mobile platform and the prototype was clear enough so that they could complete the tasks on the worksheet so that they could get a good informed opinion of the prototype which was very positive.

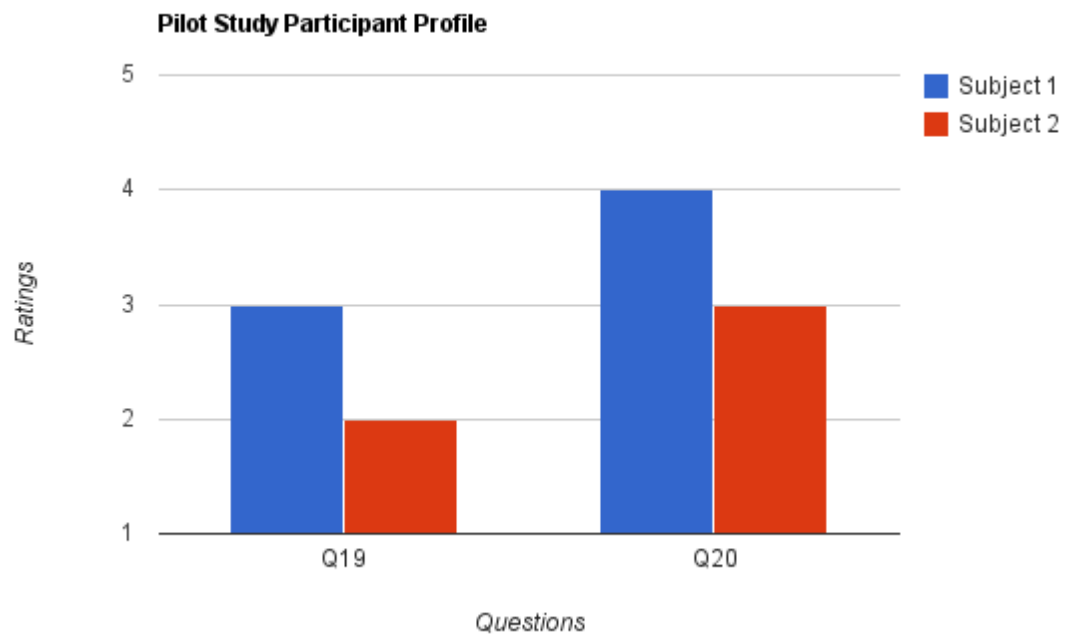


**Figure 5.4: Pilot Studies Application Analysis**

#### **5.2.1.4. Pilot Study Participant Profile**

The questions in the Participant Profile were analysed separately. The questions just ask the participants about their experience in translation and using crowdsourcing web applications. The rating responses to the questions go from “Never” (1) to the midpoint “Occasionally” (3) to “Regularly” (5). The first question in the section is question 19 , figure 5.5 shows that Subject 1 had more translation experience rating themselves a 3 where Subject 2 has less translation experience rating themselves a 2 for question 19. This means that both participants particularly Subject 1 have average levels of experience in translation. Both participants though are experienced with crowdsourcing websites with Subject 1 rating themselves a 4 and Subject 2

rating themselves a 3 . This means the participants in the pilot study had experience using crowdsourcing websites so that they have experience in using the components in crowdsourcing websites. These results show that both participants had similar experience levels in translation content and crowdsourcing websites. This gives increased confidence in the pilot study questionnaire results as the participants at least had some experience in the field of translation and crowdsourcing applications. The last question 21 which can't be shown on a graph showed that all participants in the pilot study used the Android default web browser.



**Figure 5.5: Pilot Study Participant Profile**

### 5.3 Pilot Study Feedback

The participants gave the following feedback in the feedback section of the questionnaire:

Participant	Feedback
Subject 1	<i>"Yea, I broke it on one of my jobs. Report follows stated a parse error on the translation content of the job where there was quotation marks ("")around text. Other than that no</i>

	<i>other errors in the application.”</i>
Subject 2	<i>Had to scroll too much. Turning to landscape didn't help as the table just got bigger. That's just GUI stuff though. No errors. Translation worked.”</i>

### **Other comments pilot study**

Other comments noted from the pilot participants were:

- That it would be nice to have a commenting feature similar to Facebook where the jobs could be discussed.
- There was also a bug noted in the rating system in that it allowed users to like contributions a number of times which could lead false user emergence in the application.
- Subject 1 stated that it would be useful if user names were clickable in the prototype such that all locations where user names occur could be clicked on. When a user name is clicked on it would hyperlink to a user page for the user that was clicked on . This page could display the jobs that particular user created (the user name that was clicked), the users reputation score and a small biography about their interests.

### **5.3.1 Amendments**

Based on the above participants feedback and comments the following changes were made to the prototype:

#### **Parse Error**

The parse error was fixed by using the replace function in the Java String library. This means any job source translation content or user contribution translation content given to the prototype has all occurrence of both quotation marks (“”) replaced with single quotation marks (‘’) before segmentation of the content into sentences takes place. This helped fixed the bug and this problem did not cause any issues in the main study.

#### **Rating Bug**

This rating bug was fixed by providing a MySQL table which tracks what contribution have been rated by what user in a particular job. The MySQL table was called tracklikes. When a user

rates a contribution it stores the user name, the job number where the contribution is and the unique contribution number of the content that was rated as well as changing the reputation score of the user who gave the contribution that was rated. If the user tries to rate content twice, the rating will not be increased and the user who gave the contribution will not have their reputation score increased. This helped to solve the issue raised of where users could raise their reputation score artificially by rating their contributions a numerous amount of times in the prototype.

### **Commenting**

This was added in based on comments of the pilot participants which said it would be useful to be able to comment on jobs and user contributed translations in the job contribution page. The comment section was made similar to that seen on many social networking websites where users could comment on jobs and then other users could like or dislike the comments or reply to other users comments. The data about the comments is stored in a MySQL table called comments. The comments were stored in this MySQL table with a unique comment identification number, the comment, the like and dislike rating of the comment and the user that created the comment. The comment section is located at the bottom of the job contribution page and is accessible by clicking on the comments tab in that section.

### **Clickable User Names**

This was added in when Subject 1 in the pilot study stated that they would like to be able to quickly see what jobs a particular user has created along with being able to see that users reputation score and interests in the prototype via a simple profile page. This was added in by making any location where a user name appears to be a hyperlink to that user names profile page which only display the public jobs that user has shared with the community, that users reputation score and that users interests. The jobs displayed on this page can also be clicked on to gain access to the job.

### **Fixed Formatting**

Due to complaints of both participants of having to constantly scroll up and down to see data that has to be translation, a translation area was added into the prototype near the area where users enter their translation contribution. This translation area contains data which will help to ease the life of contributors because the data that has to be translation is displayed in individual sentences in this area as well as the language that has to be translated to.

## **5.4 Main Study**

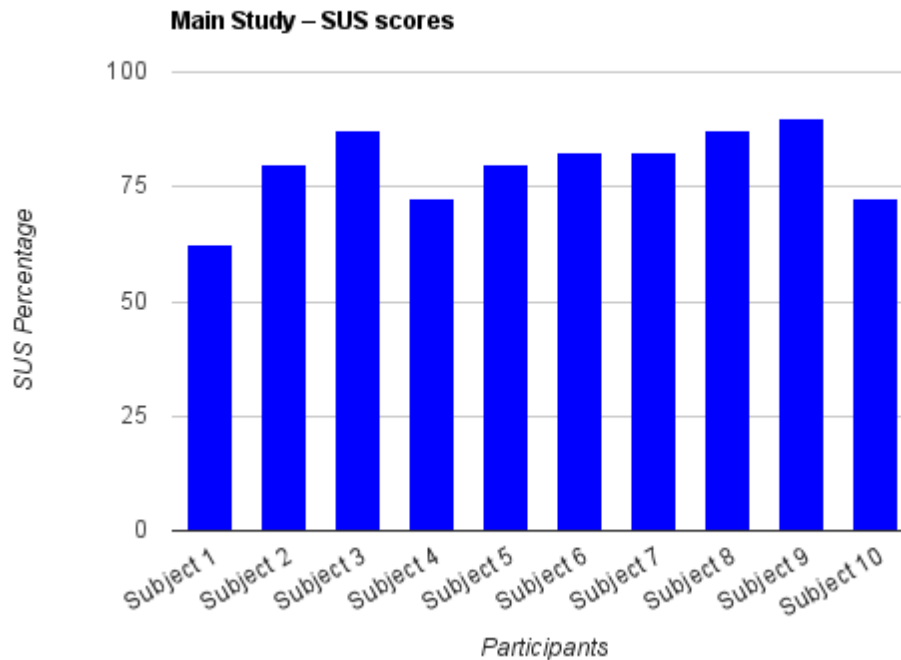
This section will discuss the main study of the evaluation with the updated prototype based on the results of the pilot study. The main study occurred over about a two week period with 10 participants who were bilingual and owned a mobile device such as a smart phone or a tablet. The testing took place privately by the individual participants and not in pairs in order for participants to give their own opinion on the prototype and not to be swayed by an observer or by other participants. The participants therefore could do the evaluation in their own time. To allow this the prototype was hosted on a web server on [eatj.com](http://eatj.com). All users were given consent forms to sign prior to engaging in the evaluation also all participants were given user names to which they were meant to create a user account to along with the worksheet of jobs they are meant to complete in the application. Once they have completed the worksheet the participants could have an informed opinion on the prototype such that the participants can answer a questionnaire on the prototype which was hosted via Google Docs. This main study will analyse these questionnaire results to see:

- How usable the prototype is via SUS questionnaire,
- How well the collaboration features are designed in the prototype,
- How well the participants perceive the concept and execution of the prototype,
- What is the users experience with translation of content and using crowd sourcing applications and
- What range of mobile devices was the prototype tested on?

### **5.4.1 Main Study Findings**

#### **5.4.1.1 System Usability Scale**

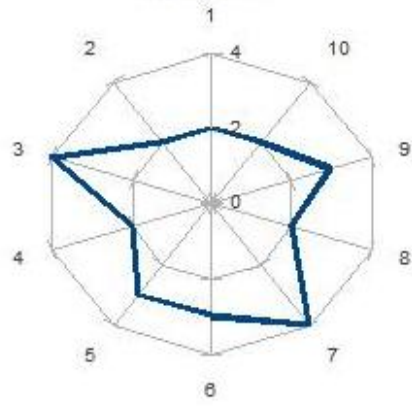
The SUS average score for the main study is 77.05% (figure 5.8). The SUS score that stands out from the participants is Subject 1. This can be clearly seen by the poor SUS rating score given by Subject 1 which is 62.5% and the misshapen star graph that Subject 1 makes in figure 5.7. Other than Subject 1's score all other participants SUS scores for the prototype are above 70 % with the second lowest being 72.5%. These SUS scores are very promising and show clear improvements in comparison to the pilot studies SUS scores with the overall average improving from 72.50% in the pilot study to 77.05% in the main study.



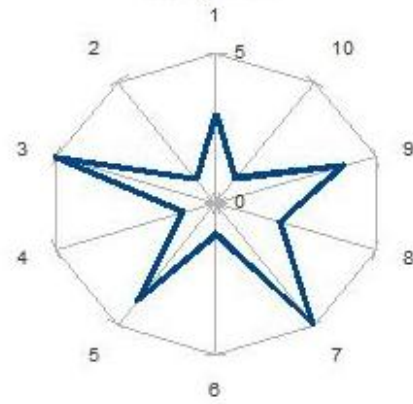
**Figure 5.6: SUS Results**

All the main study participants are shown in the radar graphs in figure 5.7. Most of the participants have well defined star shape graph which helps to indicate that these participants are satisfied with the usability of the prototype. Other participants such as Subject 1 with an irregular star shape shows a mild dissatisfaction with certain usability aspects of the prototype. This mild dissatisfaction is indicated by Subject 1 scoring questions 7 and 3 very high whilst giving average or poor ratings to the rest of the questions in the SUS questionnaire. The combination of a high average SUS score and the individual high scores of the participants (figure 5.6) indicates that the participants overall thoughts on the usability of the prototype was very positive.

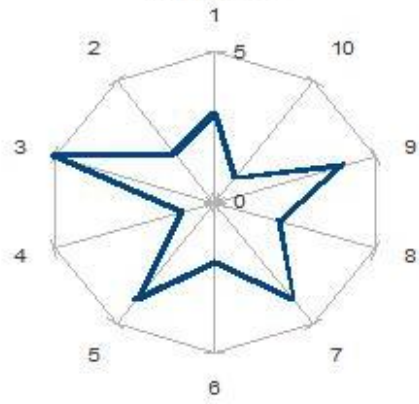
Subject 1



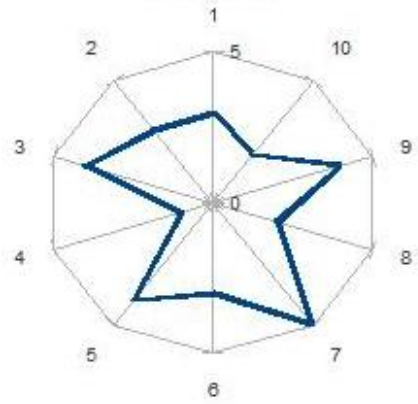
Subject 3



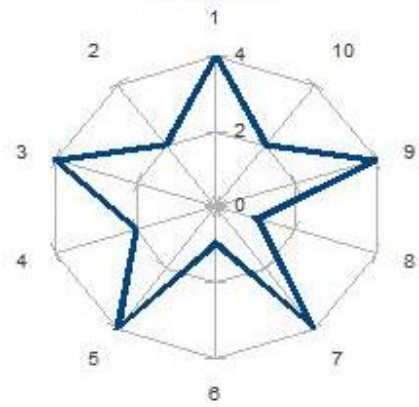
Subject 2



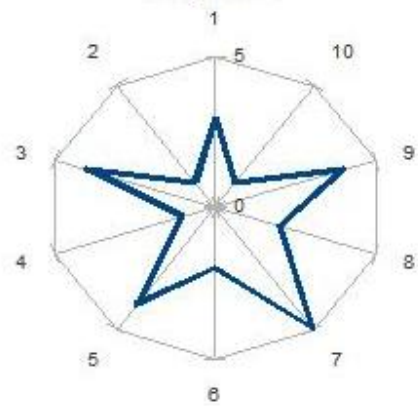
Subject 4



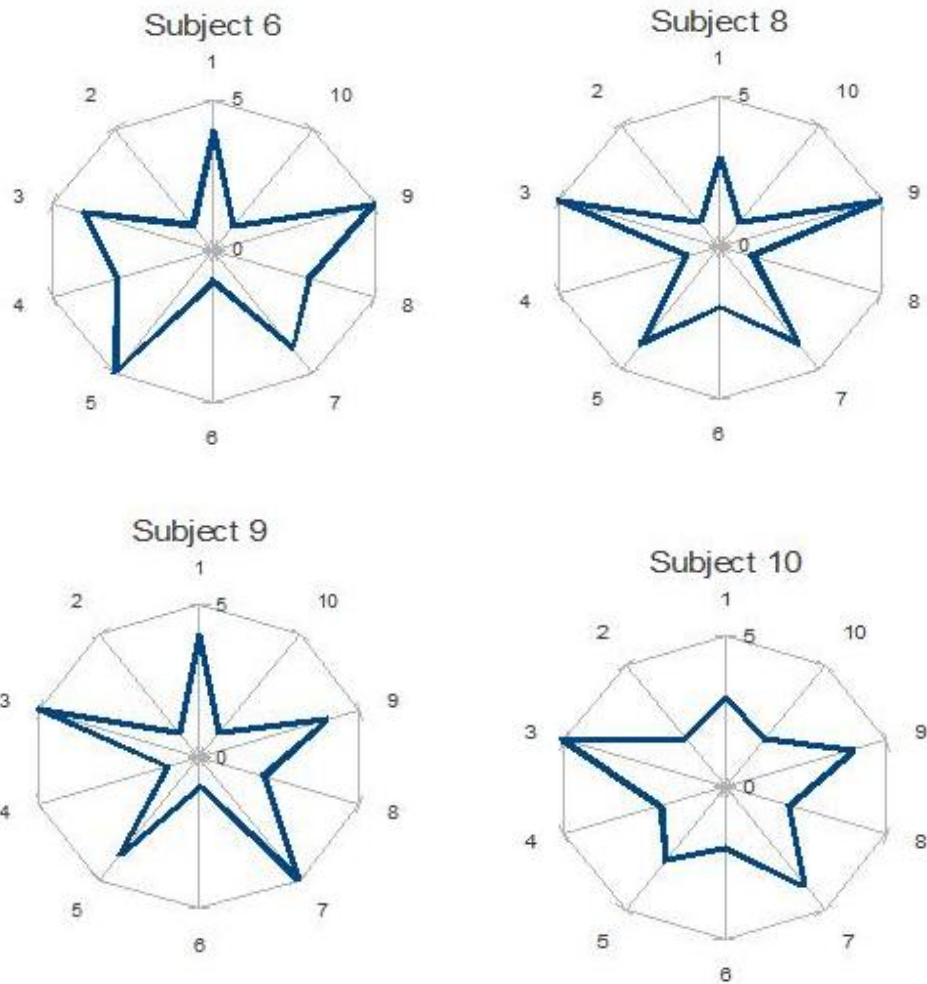
Subject 5



Subject 7

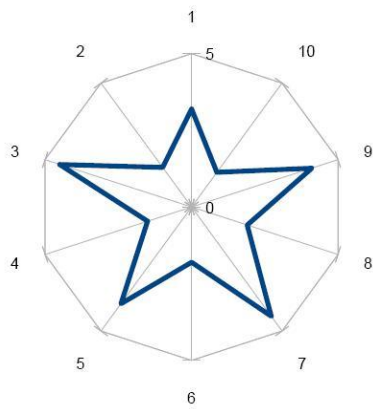






**Figure 5.7: System Usability scale results for the each test subject**

System Usability Scale: 77.05 %



**Figure 5.8: Average System Usability scale result**

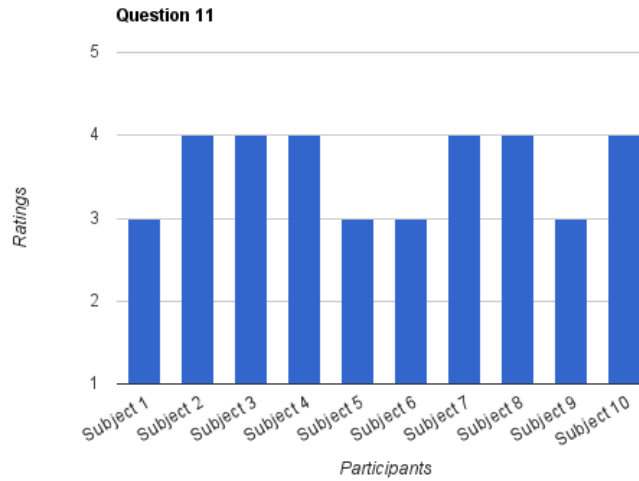
#### **5.4.1.2 Collaboration Analysis**

The second part of the questionnaire focuses on questioning how effective the design decisions were in relation to the collaboration component in the prototype. This section of the questionnaire is made up of five questions and each of these questions will be evaluated in detail in this section.

##### **Question 11 (*The prototype supports the emergence of contributor reputation*)**

The average score this question got was 3.6 with a standard deviation of 0.52 and all the participant ratings were either a 3 or 4 rating. This received the joint third highest rating score of the Collaboration Analysis questions. This shows that participants thought that the prototype handled user emergence satisfactorily through the use of a reputation score in the prototype. The users were able to assess this through looking at the reputation board in the prototype or by doing task 3 on the worksheet which was basically the participants giving contribution to the prototype. The participants would have been able to see how users reputation score increases every time they contribute positively to the community. The results could be improved in the future by having longer test period and attaining more test data such that user emergence could be more visible to participants over a sustained period of time compared to the short time and low amount of participants available in this evaluation. These results could also be improved with the addition of:

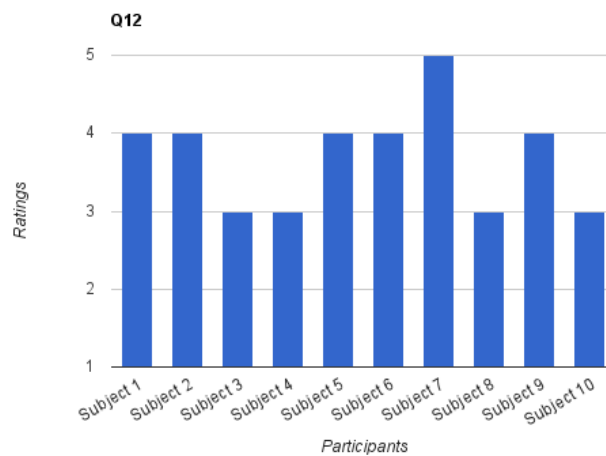
- A news feed page which tells participants who's contributing what to the community similar to Facebooks news feed page,
- The ability for users of the application to track other contributors leading to an increase in that contributors reputation score within the community and
- A ranking chart of the users in the prototype with the users with the best reputation score in the top 10 of the ranking chart.



**Figure 5.9: Question 11**

***Question 12 (The prototype supports the transparency of contributions (nature, importance))***

This question received the second highest results among the Collaboration Analysis questions receiving an average of 3.7 and a standard deviation of 0.67. The ratings are mostly between 3 and 4 similar to question 11 but Subject 5 gave a rating of 5 for this question. This shows that participants are largely satisfied with the transparency of contributions in the prototype. The users were able to form opinions on these features through the exercises that they completed on the worksheet. As discussed previously these results could be improved on by adding in additional features like news feed page showing what contributions users are making to the community.

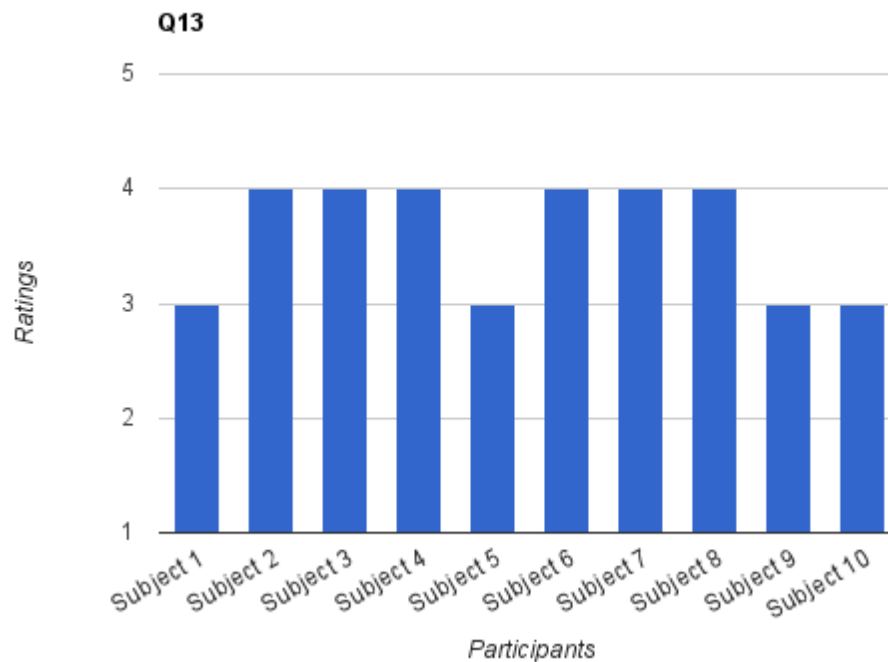


**Figure 5.10: Question 12**

**Question 13 (The prototype supports the expression of sociability)**

This question had the third highest rating along with question 11. The question had an average rating of 3.6 and a standard deviation of 0.52. Similar to question 11 the responses from the participants were mostly 3 or 4 ratings. Most of the participants ratings were 4 (6 out of 10) which indicates that the participants in large were satisfied with the sociability aspects of the prototype. The participants would have experienced the sociability aspects in the prototype by contributing comments and ratings to the application and also through the user pages. The results to this question could be improved in the future through the addition of additional sociability aspects such as:

- Users being able to add in profile pictures to user accounts,
- Adding in a news feed feature similar to Facebook,
- Adding in the ability to track/favourite users in the application and
- Giving users the ability to add in additional information about themselves such as a biography area and a place where users can write down their translation/educational experience.

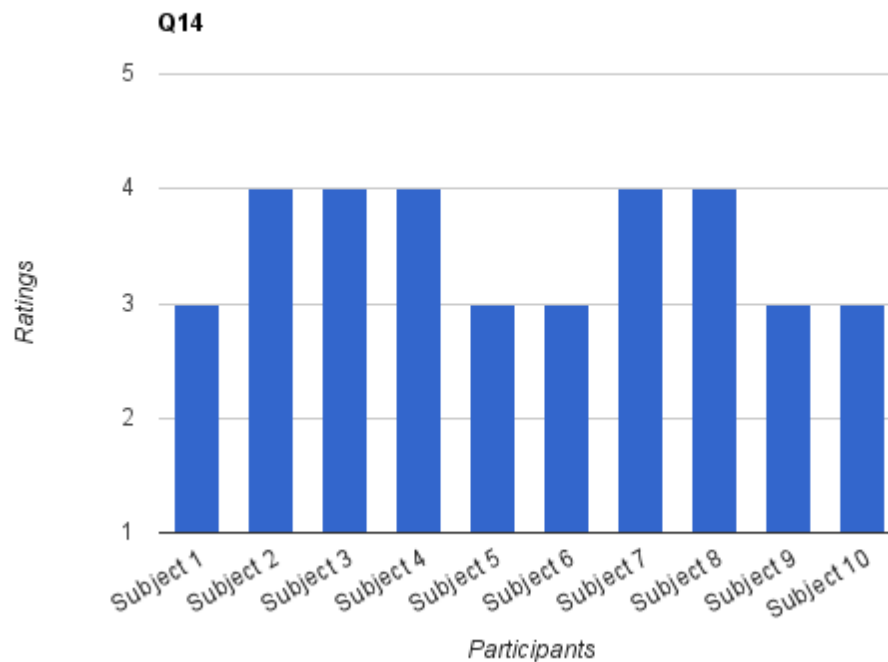


**Figure 5.11: Question 13**

**Question 14 (The prototype supports making decisions on contributions in the application (i.e. selecting best translation to publish) )**

This question was the lowest rated question in the collaboration analysis scoring an average score of 3.5 and had a standard deviation of 0.53. The rating scores were between 3 and 4 with 5 out of the 10 participants rating 4 to the question. These high ratings show that users were comfortable with decision processes in the application. The participants would have to use these decision processes whilst completing the worksheet exercise where they have to publish a translation which involves selecting the best translation out of a bunch of participants translations and also where they had to rate other users translations.

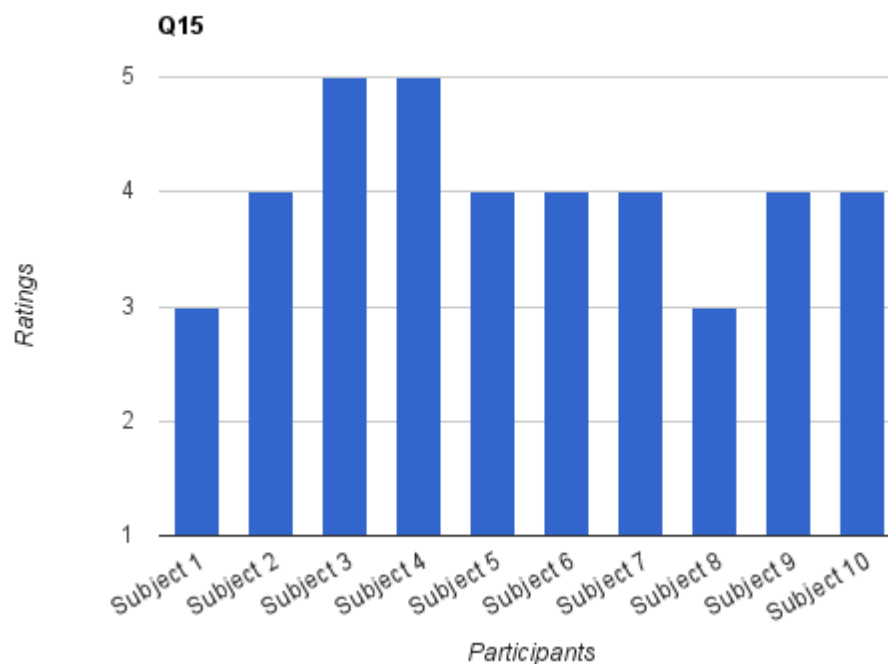
These scores could be improved in the future by allowing the decision process to be more of a key element in the application. For example allowing old jobs along with their best selected translation to be seen by the community. Instead of now in the application where the old job is only visible to the creator of the job and selecting the best translation increases the user reputation score of the person who gave the best user contributed translation. Allowing the community to see these jobs would give this component more weight in the application.



**Figure 5.12: Question 14**

**Question 15 ( The prototype supports for multiple different types of job distribution approaches (privately sharing jobs and creating custom jobs)**

This was the highest rated question of the collaboration analysis questions achieving an average rating of 4 and a standard deviation of 0.67. This shows that the participants were satisfied with the job distribution options (public or private) in the application . These features would have been tested by the user when they were doing the create a job exercise on the worksheet. The results of this question could be improved in future prototype iterations with a larger base of test participants.



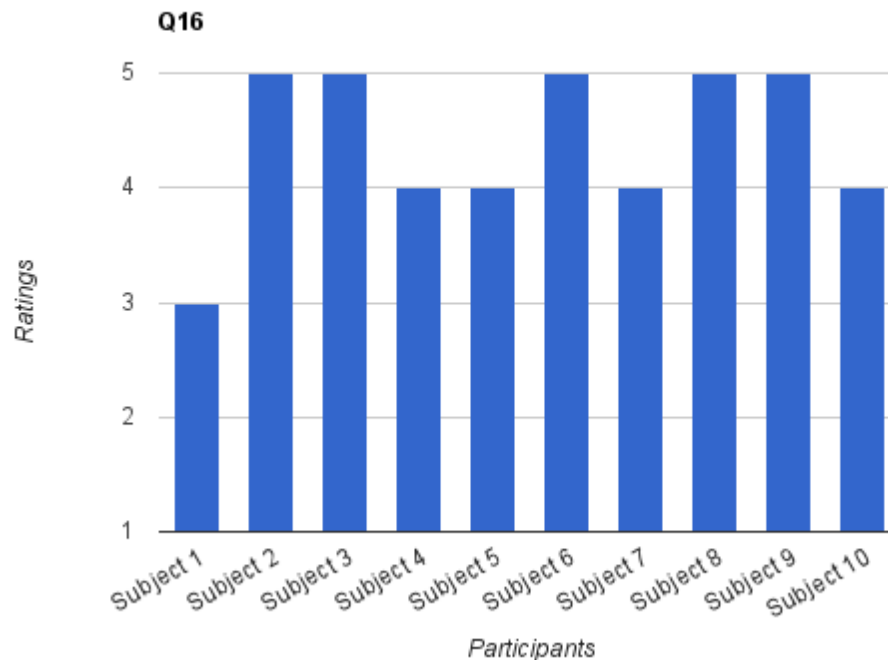
**Figure 5.13: Question 15**

**5.4.1.3 Application Analysis**

This section focuses on questioning what the participants thought of the application and did they think the application was suitable for a mobile device as well as asking the participants were they able to finish all the exercises on the worksheet (Appendix D).

***Question 16 (Do you feel that you successfully completed all the tasks on the task sheet)***

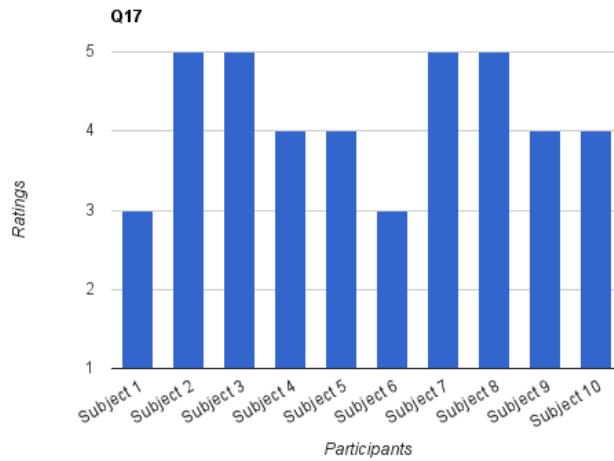
This was the highest rated question in the application analysis section of the questionnaire and it shows that most off the participants successfully completed the exercises on the worksheet (Appendix D). This helps to show how easy it is to use the main components in the prototype. This question received an overall average score of 4.4 and a standard deviation of 0.7. The majority of participants rated this question 4 or higher.



**Figure 5.14: Question 16**

***Question 17 (My overall impression of the prototype is)***

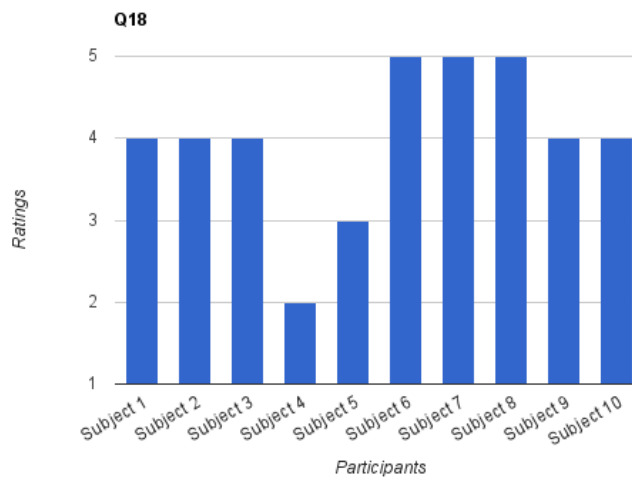
This was the second highest rated question in the Application Analysis section of the questionnaire and it got an average rating of 4.2 with a standard deviation of 0.79. This high average rating indicates that the participants in the evaluation have a positive impression of the prototype after using it.



**Figure 5.15: Question 17**

***Question 18 (Do you feel that you that this type of application is suited to a mobile device)***

This was the lowest rated of the Application Analysis section with an average of 4 and the standard deviation is 0.94. The high rating scores though indicates that the majority of participants think that this type of application is suitable for mobile device (9 out of 10 participants). The only stand out rating was by Subject 4 who gave a rating of 2 to this question.



**Figure 5.16: Question 18**



#### 5.4.1.4 Participant Profile

The questions in this section are used to get a profile on the participants' experience in using crowdsourcing applications and their experience in translating content. This section is made up of three questions and the details of the responses from the participants are detailed in the sections below.

##### *Question 19 Your proficiency and experience in translation*

This was the highest rated question in the Participants Profile and scored an average of 2.8 out of 5 and had a standard deviation of 1.03. This question was rated quite poorly apart from Subject 5 who rated the question 5 showing that this participant had the best translation experience in the group. Subject 10 also rated the question quite well also by giving a 4 rating to the question. The other eight participants rated the question quite lowly with three participants rating their translation experience to be average at 3 and five rating their translation experience quite lowly at 2. This shows that the prototype had mostly a good mixture of experienced translators with a rating of 3 or higher and not so experienced translators involved in the evaluation of the prototype. In the future it would be more advantageous to have mostly experienced translators involved in the testing of the application as they would have more experience proof reading machine translations and would show more of an interest in a mobile crowdsourcing translation tool.

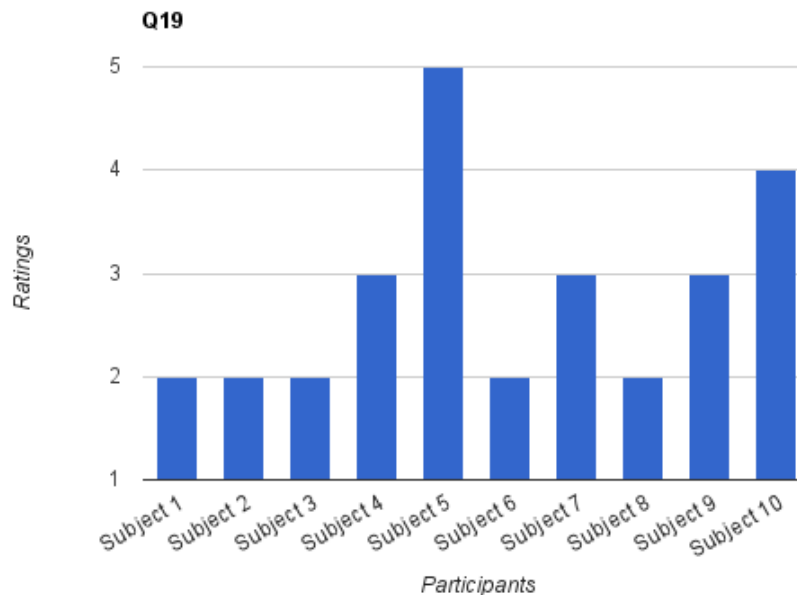
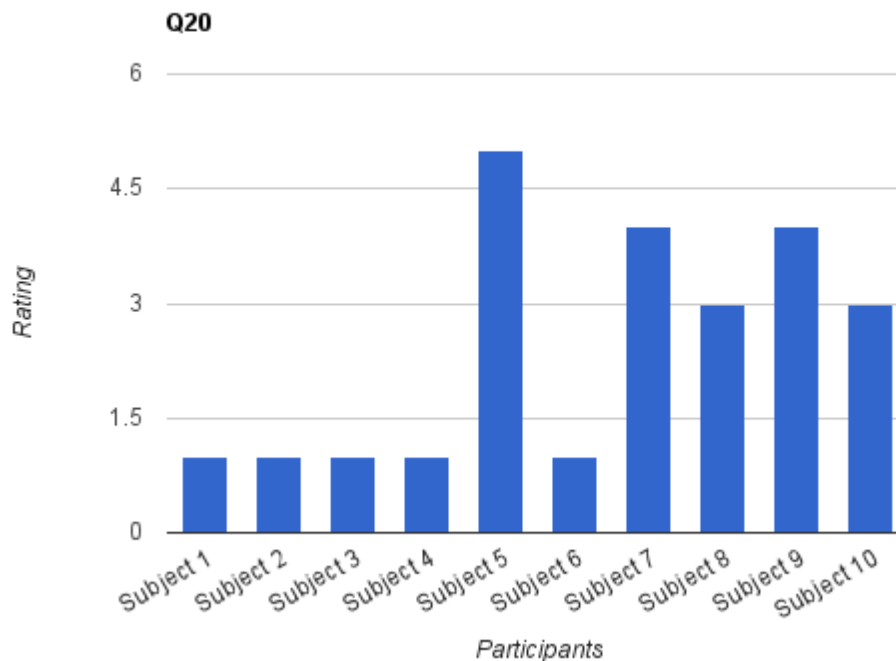


Figure 5.17: Question 19

***Question 20 Your level of contribution to online crowdsourcing sites***

This was the lowest rated question in the Participants Profile section getting only an average score of 2.4 out of 5 and a standard deviation of 1.6. The highest rated participant again is Subject 5 who gave a rating of 5 and this shows that the Subject 5 has good experience in translating and using crowdsourcing web sites. Similar to question 19 five of the participants gave a rating of 3 or higher to the question indicating that five of the participants had some experience of crowdsourcing concepts and issues. The low ratings given by the other five participants indicates that these participants have little familiarity with crowdsourcing concepts used in this prototype and therefore little familiarity with some of the issues involved in this type of application. This question highlights that there was a good mixture of novice and expert users of crowdsourcing sites and overall these users were impressed with usability of the prototype and its collaboration features.



**Figure 5.18: Question 20**

***Question 21 What phone or web Browser did you use?***

The participants largely used the prototype on the Android handset with 6 out of the 10 using the default Android browser (Android) on their handset to do the evaluation. The 2nd largest was

the iPhone with 3 out of the 10 participants using its default browser (Safari) to do the evaluation. Finally, the least used handset in the evaluation was the Windows mobile handset which only 1 out of the 10 participants using the default web browser (Internet Explorer Mobile) to do the evaluation. This shows that a good mixture of smart phone handsets were used in the evaluation of the prototype and this indicates with the SUS average score of 77.05% that the prototype has high usability across the board on smart phone handsets. In future studies perhaps a more wide variety mobile devices and mobile web browser could be tested on such as Opera Mobile, Firefox Mobile, Blackberry browser etc.

<b>Participants</b>	<b>Web Browser</b>
Subject 1	Android Browser
Subject 2	Android Browser
Subject 3	Safari
Subject 4	Android Browser
Subject 5	Internet Explorer Mobile
Subject 6	Android Browser
Subject 7	Safari
Subject 8	Android Browser
Subject 9	Android Browser
Subject 10	Safari

**Figure 5.19: Mobile browsers used in the evaluation**

#### **5.4.2 Feedback**

In the feedback section of the questionnaire, 4 participants wrote some feedback comments about the prototype. The only feedback that was repeated was that a participant felt like they had to do a bit of scrolling to see the text that had to be translated. The positive aspect of this was

that none of the other errors from the pilot study were repeated again and therefore these major issues did not disrupt the overall course of the main studies results.

Two of the participants commented on web page layout saying:

***“Tweak the layout so the translation text is more visible whilst making contribution”***

***“Difficult to see the whole page with the translation work and thus get an idea of what was on the page.”***

One participants wanted more languages available in the prototype:

***“Users may want to add new languages that they speak to the language list”***

One participant wanted better sociability aspects to be available in the prototype:

***“Improve social networking aspect with friends, messaging and the such”***

The feedback about extra languages is outside the scope of this dissertation but will be handled in future work done on this application. The layout issue will be investigated again and a newer easier to use interface will be created using testers to test out the interface prior to the next large evaluation. The improving of the social aspect will also be investigated in future work.

### **5.5 Evaluation summary**

A total of 12 participants took part in the evaluation and had no prior knowledge of the research prior to taking part in the evaluation. The participants had to be bilingual and own a mobile device like a smart phone or a tablet. The evaluation consisted of a worksheet containing five exercises which had to be completed using a mobile device followed by a questionnaire. The questionnaire asked the participants what they thought of the usability and collaboration features in the prototype as well asking them about the participants relevant experience in translation and with crowdsourcing web applications.

The first two participants took part in a pilot study to help in ironing out any issues in the prototype prior to doing the main study. The pilot participants did their testing under supervised condition where they could give back feedback in both the questionnaire and through verbal feedback. Once the pilot study was completed and the issues that arose were fixed then the main study could take place. The main study consisted of ten participants who were given an exercise

sheet to complete along with a questionnaire. These participants in the main study worked privately without any input from outside sources.

The data retrieved indicated that the participants were generally pleased with the usability the prototype giving a high average SUS score of 77.05% and the participants were pleased with the collaboration features in the prototype also. The one thing that came through in feedback was that the layout of job content in relation to translation content could be improved as two participants stated their displeasure with it and another participant would like improved sociability aspects in future prototypes. In future evaluations, more time and a larger community of professional or semi-professional translators would be useful in getting a more critical and insightful comments on future iterations of the prototype . To conclude, the participants that took part in the evaluation were satisfied with the prototypes usability and collaboration features.

# Chapter 6

## Conclusion

This final section will discuss what possible future work could be done to build upon the work already completed along with concluding thoughts about the dissertation.

### 6.1 Conclusion

In conclusion, the dissertation achieved its goal of building a usable mobile crowdsourcing translation application that could be used in both rich and poor economic regions using semantic web technologies. The reason behind the goal was that most crowdsourced translation systems on the market are designed and targeted at desktop based browsers inspired by the design of professional computer assisted translation tools. This impedes a large population of potential translators in poor economic regions where access to a PC is a rarity. In these poorer regions the majority of users do their personal and business computing through mobile devices. Similarly in some regions even where economic development is strong some users still do their personal computing through their mobile device. This leads to the obvious conclusion of building a mobile crowdsourcing translation application. This solution is particularly advantageous because a large amount of the world's population carry their phones with them the entire day giving them the opportunity to contribute at any time to the translation process.

The use of RDF technologies played a key role in storing localization data in the application. RDF was chosen to store the localization data over RDBMS because building up linked relationships between translation jobs which use similar languages or jobs in the same category can be done more easily and organically in the evolving RDF graph of data compared to the more structure RDBMS environment. These language resources stored in the RDF triple store will help in the training of domain-specific data driven language technologies (Statistical Machine Translation) and also help in long-tail translation projects both of which will be part of the future work in this project.

In the evaluation, data was retrieved through usage studies of the prototype which showed that participants were generally pleased with the usability of the prototype giving a high average

SUS score of 77.05% and the participants were also satisfied with the collaboration features in the prototype rating them highly. The only negative of the evaluation was that the community was not big enough with the main study only taking place using ten participants. In future evaluations, more time and a larger community of professional or semi-professional translators would be useful in getting a more critical and insightful comments on future iterations of the prototype.

## **6.2 Future Work**

To conclude, the prototype and positive results gotten through the research in this dissertation can be built upon and improved in future iterations by:

- Making more languages available in the prototype by allowing right to left scripted languages to be included in the application such as Chinese, Japanese, and Hindi etc. The application in the future iteration should also include more left to right scripted languages such as Danish, Finnish, and Irish etc.
- The ability to correlate the users contributions to achieve the best possible translation from the crowd as discussed by Sheng et al (2008).
- Improve the layout of translation content in the application because participants in the evaluation said they had to do a lot of scrolling to see translation content. This layout issue will be investigated and a newer easier to use interface will be created using testers to test out the multiple interfaces prior to the next large evaluation.
- A more in-depth evaluation with a sustained online community of professional or semi-professional translators for more insightful and conclusive results.
- Using the localisation data in the RDF triple store to help in the training of domain-specific statistical machine translation or to aid in long-tail translation projects.
- Better segmentation of translation content to fix a noted error where words such as St.Patrick are segmented incorrectly by thinking that "." is the end of one sentence and that Patrick is the start of another sentence. This can be fixed easily by the use of

defining segmentation rules for segmenting sentences in the application. A Java library which will be able to support this text segmentation is Apache Lucene [5]. Apache Lucene is a fully featured text engine library which is written in Java and supports text segmentation in its *analysis.standard* package which is built using JFlex (lexical analyser).

- A more fleshed out social networking aspect to include features such as:
  - Users being able to customize their profile page by a profile pictures and the ability for users to add more information about themselves such as a biography, languages that they speak, educational history, translation experience etc.
  - Adding in a news feed feature similar to Facebook which will display other users contributions and ratings within the community.
  - Adding in the ability to track/favourite users in the application.
  - Adding the ability to send messages to other users in the application.



# Appendix A

## MySQL

```
--  
-- Database: `killer7`  
--  
CREATE TABLE IF NOT EXISTS `comments` (  
  `job_id` int(11) default NULL,  
  `comment` varchar(1000) default NULL,  
  `liking` varchar(100) default NULL,  
  `disliking` varchar(100) default NULL,  
  `username` varchar(100) default NULL,  
  KEY `job_id` (`job_id`),  
  KEY `username` (`username`)  
) ENGINE=MyISAM DEFAULT CHARSET=latin1;  
  
CREATE TABLE IF NOT EXISTS `dist` (  
  `job_id` int(11) default NULL,  
  `username` varchar(100) default NULL,  
  KEY `job_id` (`job_id`),  
  KEY `username` (`username`)  
) ENGINE=MyISAM DEFAULT CHARSET=latin1;  
  
CREATE TABLE IF NOT EXISTS `interest` (  
  `username` varchar(100) default NULL,  
  `interest` varchar(100) default NULL,  
  KEY `username` (`username`)  
) ENGINE=MyISAM DEFAULT CHARSET=latin1;  
  
CREATE TABLE IF NOT EXISTS `Translation` (  
  `comment_id` int(11) NOT NULL auto_increment,  
  `job_id` int(11) default NULL,
```

```

`contribution` mediumtext,
`liking` varchar(100) default NULL,
`disliking` varchar(100) default NULL,
`username` varchar(100) default NULL,
PRIMARY KEY (`comment_id`),
KEY `job_id` (`job_id`),
KEY `username` (`username`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=7 ;

```

```

CREATE TABLE IF NOT EXISTS `jobs` (
`job_id` int(11) NOT NULL auto_increment,
`username` varchar(100) default NULL,
`jobname` varchar(100) default NULL,
`jobdescription` varchar(10000) default NULL,
`category` varchar(50) default NULL,
`datecreated` varchar(100) default NULL,
`datefinished` varchar(100) default NULL,
`published` tinyint(1) default NULL,
`dist` varchar(100) default NULL,
PRIMARY KEY (`job_id`),
KEY `username` (`username`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=62 ;

```

```

CREATE TABLE IF NOT EXISTS `tracklikes` (
`job_id` int(11) default NULL,
`refname` varchar(100) default NULL,
`username` varchar(100) default NULL,
KEY `username` (`username`)
) ENGINE=MyISAM DEFAULT CHARSET=latin1;

```

```

CREATE TABLE IF NOT EXISTS `users` (
`id` int(11) NOT NULL auto_increment,
`firstname` varchar(100) NOT NULL,

```

```
`lastname` varchar(100) NOT NULL,  
`username` varchar(100) NOT NULL,  
`email` varchar(100) NOT NULL,  
`password` varchar(100) NOT NULL,  
`rating` int(11) NOT NULL,  
PRIMARY KEY (`username`,`id`),  
UNIQUE KEY `id` (`id`)  
) ENGINE=MyISAM DEFAULT CHARSET=latin1 AUTO_INCREMENT=17 ;
```

# Appendix B

## Questionnaire

**Please note:**

Each question is optional. Feel free to omit a response to any question; however the researcher would be grateful if all questions are responded to. Please do not name third parties in any open text field of the questionnaire. Any such replies will be anonymised. In the extremely unlikely event that illicit activity is reported I will be obliged to report it to appropriate authorities.

### 1. System Usability Scale

1. I think that I would like to use this system frequently

<b>Strongly disagree</b>					<b>Strongly agree</b>
1	2	3	4	5	

2. I found the system unnecessarily complex

<b>Strongly disagree</b>					<b>Strongly agree</b>
1	2	3	4	5	

3. I thought the system was easy to use

<b>Strongly disagree</b>					<b>Strongly agree</b>
1	2	3	4	5	

4. I think that I would need the support of a technical person to be able to use this system

<b>Strongly disagree</b>					<b>Strongly agree</b>
1	2	3	4	5	

5. I found the various functions in this system were well integrated

<b>Strongly disagree</b>					<b>Strongly agree</b>
1	2	3	4	5	

6. I thought there was too much inconsistency in this system

**Strongly disagree**

**1            2            3            4            5**

**Strongly agree**

7. I would imagine that most people would learn to use this system very quickly

**Strongly disagree**

**1            2            3            4            5**

**Strongly agree**

8. I found the system very cumbersome to use

**Strongly disagree**

**1            2            3            4            5**

**Strongly agree**

9. I felt very confident using the system

**Strongly disagree**

**1            2            3            4            5**

**Strongly agree**

10. I needed to learn a lot of things before I could get going with this system

**Strongly disagree**

**1            2            3            4            5**

**Strongly agree**

## **2. Sociability Support**

11. The prototype supports the emergence of contributor reputation

**Strongly disagree**

**1            2            3            4            5**

**Strongly agree**

12. The prototype supports the transparency of contributions (volume, nature, importance)

**Strongly disagree**

**1            2            3            4            5**

**Strongly agree**

13. The prototype supports the expression of sociability (etiquette and social rules)

**Strongly disagree**

**1            2            3            4            5**

**Strongly agree**

14. The prototype supports making decisions on contributions in the application (i.e. selecting best translation to publish)

**Strongly disagree**

**1**

**2**

**3**

**4**

**5**

**Strongly agree**

15 . The prototype supports for multiple different types of distribution approaches and changes in the job creation work flow (privately sharing jobs and creating custom jobs)

**Strongly disagree**

**1**

**2**

**3**

**4**

**5**

**Strongly agree**

### 3. Application Analysis

16. Do you feel that you successfully completed all the tasks on the task sheet?

**Strongly disagree**

**1**

**2**

**3**

**4**

**5**

**Strongly agree**

17. My overall impression of the prototype is:

**Strongly Negative**

**1**

**2**

**3**

**4**

**5**

**Strongly Positive**

18. Do you feel that you that this type of application is suited to a mobile device:

**Strongly disagree**

**1**

**2**

**3**

**4**

**5**

**Strongly agree**

### 4. Participant Profile

19. Your proficiency and experience in translation

**Never**

**1**

**2**

**Occasionally**

**3**

**4**

**Regularly**

**5**

20. Your level of contribution to online crowd sourcing sites (boards.ie, Wikipedia etc.)

**Never**

**1**

**2**

**Occasionally**

**3**

**4**

**Regularly**

**5**

21. What phone or web Browser did you use?

.....

**Feedback**

If you have any comments on this study, please write them here:

.....

.....

.....

.....

.....

# Appendix C

## Data Collected

### C.1 Pilot Studies Results

#### SUS Pilot Study Results

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Subject 1	3	2	4	1	4	2	4	2	4	2
Subject 2	3	2	3	2	4	2	5	1	3	3

#### Sociability Support Pilot Study Results

	Q11	Q12	Q13	Q14	Q15
Subject 1	4	3	4	3	4
Subject 2	4	4	4	4	5

#### Application Analysis Pilot Study Results

	Q16	Q17	Q18
Subject 1	5	4	5
Subject 2	5	4	4

#### Participant Profile Pilot Study Results

	Q19	Q20
Subject 1	2	3
Subject 2	2	2



## C.2 Main Studies Results

### SUS Results

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Subject 1	2	2	4	2	3	3	4	2	3	2
Subject 2	3	2	5	1	4	2	4	2	4	1
Subject 3	3	1	5	1	4	1	5	2	4	1
Subject 4	3	3	4	1	4	3	5	2	4	2
Subject 5	4	2	4	2	4	1	4	1	4	2
Subject 6	4	1	4	3	5	1	4	3	5	1
Subject 7	3	1	4	1	4	2	5	2	4	1
Subject 8	3	1	5	1	4	2	4	1	5	1
Subject 9	4	1	5	1	4	1	5	2	4	1
Subject 10	3	2	5	2	3	2	4	2	4	2

### Sociability Support Results

	Q11	Q12	Q13	Q14	Q15
Subject 1	3	4	3	3	3
Subject 2	4	4	4	4	4
Subject 3	4	3	4	4	5
Subject 4	4	3	4	4	5
Subject 5	3	4	3	3	4

Subject 6	3	4	4	3	4
Subject 7	4	5	4	4	4
Subject 8	4	3	4	4	3
Subject 9	3	4	3	3	4
Subject 10	4	3	3	3	4

### Application Analysis

	Q16	Q17	Q18
Subject 1	3	3	4
Subject 2	5	5	4
Subject 3	5	5	4
Subject 4	4	4	2
Subject 5	4	4	3
Subject 6	5	3	5
Subject 7	4	5	5
Subject 8	5	5	5
Subject 9	5	4	4
Subject 10	4	4	4

**Participant Profile**

	<b>Q19</b>	<b>Q20</b>
Subject 1	2	1
Subject 2	2	1
Subject 3	2	1
Subject 4	3	1
Subject 5	5	5
Subject 6	2	1
Subject 7	3	4
Subject 8	2	3
Subject 9	3	4
Subject 10	4	3

# Appendix D

## Worksheet

Please use a smart phone for the evaluation. Please do the exercises first and then do the questionnaire. If you want any more information don't hesitate to email (finnle@tcd.ie)

### Step 1

Exercises to be completed in the prototype:

- Create a user account - please use the user name that has been supplied to you
- Create a translation job
- Contribute translation content to a translation job
- Rate other users contributions
- Comment on a translation job
- Publish a translation job

### Step 2

Complete the questionnaire provided here:

<https://spreadsheets.google.com/a/tcd.ie/spreadsheet/viewform?formkey=dGhuRGFMZDZRLXpMSHJTUWZDSmVJMkE6MQ>

Thanks,

Leroy Finn

# Bibliography

- [1] Alchemy Software, 2011. *Alchemy Software Development* [online] (Updated 05 April 2011) Available at: <<http://www.alchemysoftware.ie/>> [Accessed 29 August 2011].
- [2] Alt, F. Kramer, U. Nawaz, Z. Schmidt, A. Shirazi, A.S., 2010. *Location-based crowdsourcing: extending crowdsourcing to the real world*, In Proceedings of NordiCHI. 2010, pp.13-22.
- [3] Amazon Mechanical Turk, 2011, *Amazon Mechanical Turk* [online] Available at: <<https://www.mturk.com/mturk/welcome>> [Accessed 25 April 2011].
- [4] Android developers, 2011, *Android developers* [online] Available at: <<http://developer.android.com/index.html>> [Accessed 25 April 2011].
- [5] Apache Lucene, 2011. *Apache Lucene - Overview* [online] (Updated 28 July 2011) Available at: <<http://lucene.apache.org/java/docs/index.html>> [Accessed 29 August 2011].
- [6] Apache Tomcat, 2011. *Apache Tomcat* [online] (Updated 25 August 2011), Available at: <<http://tomcat.apache.org/>> [Accessed 29 August 2011].
- [7] Apple, 2011, *iOS Developer Program* [online] Available at: <<http://developer.apple.com/programs/ios/>> [Accessed 25 April 2011].
- [8] Architecture Resources, 2001. *Functional Requirements and Use Case* [online] Available at: <[http://www.bredemeyer.com/pdf\\_files/functreq.pdf](http://www.bredemeyer.com/pdf_files/functreq.pdf)> [Accessed 29 August 2011].
- [9] Asia Online, 2011. *Asia Online* [online] Available at: <<http://www.asiaonline.net/portal.aspx>> [Accessed 29 August 2011].
- [10] Berners-Lee, T. Hendler, J. Lassila, O., 2008. The Semantic Web, *Scientific American Magazine* [online] Available at: <<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>> [Accessed 29 August 2011].
- [11] Bioontology, 2009. *Comparison of Triple Stores* [online] Available at: <[http://www.bioontology.org/wiki/images/6/6a/Triple\\_Stores.pdf](http://www.bioontology.org/wiki/images/6/6a/Triple_Stores.pdf)> [Accessed 29 August 2011].
- [12] Bizer, C. Schultz, A., 2009. *The Berlin SPARQL Benchmark*. International Journal on Semantic Web and Information System, 5(2), pp.1-24.
- [13] Centre for Next Generation Localisation, 2010, *Crowdsourcing vs. Machine Translation* [online] Available at: <<http://www.cngl.ie/drupal/sites/default/files/papers3/Crowdsourcing.pdf>> [Accessed 25 April 2011].
- [14] Centre for Next Generation Localisation, 2011, *Research Overview* [online] Available at: <<http://www.cngl.ie/research.html>> [Accessed 25 April 2011].

- [15] Chen, X. Zhang, S. Zhang, S. Wu, S., 2010. Second International Conference on Future Networks, 2010. ICFN '10, pp. 88-92.
- [16] Chinnery, G.M., 2006. *EMERGING TECHNOLOGIES Going to the MALL: Mobile Assisted Language Learning*, Language Learning & Technology Vol 10, No.1, pp.9-16.
- [17] Curran, S. Feeney, K. Schaler, R. Lewis, D. , 2009. *The management of crowdsourcing in business processes*, Integrated Network Management-Workshops, pp.77 – 78.
- [18] Dawid, A.P. Skene, A.M., 1979. *Maximum Likelihood Estimation of Observer Error-Rates Using the EM Algorithm*, Applied Statistics, Vol 28, No1, pp 20-28.
- [19] Désilets, A., 2011. *Wanted: Best Practices for Collaborative Translation* [online] Available at: <<http://www.translationautomation.com/best-practices/wanted-best-practices-in-collaborative-translation.html>> [Accessed 25 April 2011].
- [20] Eagle, N., 2009. *txteagle: Mobile Crowdsourcing*, Lecture Notes in Computer Science, Volume 5623/2009, pp.447-456.
- [21] Evans, C. Segaran, T. Taylor, J., 2009. *Programming the Semantic Web*. 1st edition. O'Reilly Media Inc.
- [22] Farid, M., 1990. *Software engineering globalization and localization*, IEEE International Conference on Systems, Man and Cybernetics, pp.491-495
- [23] Google Code, 2011. *microsoft-translator-java-api* [online] (Updated 29 August 2011) Available at: <<http://code.google.com/p/microsoft-translator-java-api/>> [Accessed 29 August 2011].
- [24] Greenshpan, O. Kumara, S. Vukovic, M., 2010. *Ubiquitous crowdsourcing*, Ubicomp '10: Ubiquitous Computing , pp.523-525.
- [25] Hilton, B.N., 2007. *Emerging spatial information systems and applications*. Idea Group Inc, PP. 146.
- [26] Hutchins, W.J., 2000. *Early Years in Machine Translation: Memoirs and Biographies of Pioneers*, Amsterdam Studies in the Theory and History of Linguistic Science Series III: Studies in the History of the Language Sciences.
- [27] IBM, 2010. *Cloud services for your virtual infrastructure* [online] (Updated 10 January 2010), Available at: <[http://www.ibm.com/developerworks/opensource/library/os-cloud-virtual1/index.html?S\\_TACT=105AGX01&S\\_CMP=HP](http://www.ibm.com/developerworks/opensource/library/os-cloud-virtual1/index.html?S_TACT=105AGX01&S_CMP=HP)> [Accessed 25 April 2011].
- [28] Kasparek, C., 1983, *The Translator's Endless Toil*, *The Polish Review*, vol. XXVIII, no. 2, 1983, pp. 84-87.

- [29] Language-translation-help,2011. *Crowdsourcing Translations* [online] (Updated 27 June 2011) Available at:<<http://www.language-translation-help.com/crowdsourcing.html>> [Accessed 29 August 2011].
- [30] Learnthemobileweb, 2009. *Mobile META-Tags* [online] (Updated 3 July 2009), Available at:<<http://learnthemobileweb.com/2009/07/mobile-meta-tags/>> [Accessed 29 August 2011].
- [31] Lee, L. Pang, B., 2004. *A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts*, Proceedings of the ACL, pp.271-278.
- [32] Lewis, D., 2011. *Linked Language and Localisation Data*, CNGL Internal Document.
- [33] Lionbridge Translation and Localization,2010. *How Will Crowdsourcing Change the Translation Industry?* [online] (Updated 6 May 2010 ) , Available at:<<http://blog.lionbridge.com/translation-and-localization/bid/34217/How-Will-Crowdsourcing-Change-the-Translation-Industry>> [Accessed 25 April 2011].
- [34] Logic Wizards, 2011. *Cloud Comparisons: Apples to Apples* [online] (Updated 2 February 2011), Available at: <<http://www.logicwizards.net/2011/02/02/cloud-comparisons-apples-to-apples/>> [Accessed 25 April 2011].
- [35] Making Light, 2007. *From correspondence: Top this* [online] (Updated 13 August 2007), Available at: <<http://nielsenhayden.com/makinglight/archives/009274.html> > [Accessed 25 April 2011].
- [36] McArthur, T., 1992. *The Oxford Companion to the English Language*, ed.1992, pp. 1,051–54.
- [37] Mobile Translator, 2011. *Mobile translator - free online translation and dictionary*, [online] Available at:<<http://www.mobile-translator.net/> > [Accessed 25 April 2011].
- [38] Morimoto, T. Kurematsu, A. ,1993.*Automatic Speech Translation at ATR*,MT Summit IV,pp.83-96.
- [39] Multilingual Web.W3C Workshop Report:*The Multilingual Web: Where Are We?*[online] (Updated 1 November 2010), Available at:<<http://www.multilingualweb.eu/documents/madrid-workshop/madrid-workshop-report> > [Accessed 25 April 2011].
- [40] Munro, R., 2010. *Crowdsourced translation for emergency response in Haiti: the global collaboration of local knowledge*, Collaborative Translation: technology, crowdsourcing and the translator perspective, AMTA 2010.
- [41]NC3RS, 2006. *Why do a pilot study?* [online] Available at:<<http://www.nc3rs.org.uk/downloaddoc.asp?id=400>> [Accessed 29 August 2011].

- [42] NextGenLocalisation, 2011. *Service Oriented Localisation Architecture Solution (SOLAS)*. [video online] Available at: <<http://www.youtube.com/watch?v=KPdxZwRxfA4>> [Accessed 29 August 2011].
- [43] Oasis, 2008. *XLIFF Version 1.2* [online] (Updated 1 February 2010) Available at: <<http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html>> [Accessed 29 August 2011].
- [44] Oracle, 2011. *Extending ADF Mobile Client Applications with Java* [online] (Updated 24 March 2011) Available at: <[http://download.oracle.com/docs/cd/E17904\\_01/doc.11111/e14826/ext\\_java.htm](http://download.oracle.com/docs/cd/E17904_01/doc.11111/e14826/ext_java.htm)> [Accessed 29 August 2011].
- [45] Open Source, 2010. *Love, hate, and the Wikipedia contributor culture problem* [online] (Updated 8 March 2010), Available at: <<http://opensource.com/business/10/3/love-hate-and-wikipedia-contributor-culture-problem>> [Accessed 25 April 2011]
- [46] Perishablepress, 2009. *The 5 minute css mobile makeover* [online] (Updated 2 August 2009) Available at: <<http://perishablepress.com/press/2009/08/02/the-5-minute-css-mobile-makeover/>> [Accessed 29 August 2011].
- [47] Practicalecommerce, 2007. *Preparing Your Website For Mobile Devices* [online] (Updated 15 January 2007) Available at: <<http://www.practicalecommerce.com/articles/395-Preparing-Your-Website-For-Mobile-Devices>> [Accessed 29 August 2011].
- [48] Richardson, J. Romaine, M., 2009. *State of the Translation Industry 2009*, myGengo, Inc., Available at: <<http://mygengo.com/report/translation-industry-2009>> [Accessed 25 April 2011].
- [49] Seabreezecomputers, 2011. *Mobile CSS* (Updated 28 May 2011), Available at: <<http://www.seabreezecomputers.com/tips/mobile-css.htm>> [Accessed 29 August 2011].
- [50] Semantic Web, 2010. *When to use triple store?* [online] (Updated 14 July 2010) Available at: <<http://answers.semanticweb.com/questions/1312/when-to-use-triple-store>> [Accessed 29 August 2011].
- [51] Semantic Web, 2011. *Semantic Web Main Page* [online] (Updated 2 August 2011) Available at: <[http://semanticweb.org/wiki/Main\\_Page](http://semanticweb.org/wiki/Main_Page)> [Accessed 29 August 2011].
- [52] Sesame, 2009. *What is Sesame?* [online] (Updated 19 February 2009), Available at: <<http://www.openrdf.org/doc/sesame/users/ch01.html>> [Accessed 29 August 2011].
- [53] Shannon, C. E. and Weaver, W., 1963. *The Mathematical Theory of Communications*, Illinois: The University of Illinois Press, Ch. 1, pp. 3-28.
- [54] Sharma, A., 2010. *Crowdsourcing Critical Success Factor Model: Strategies to harness the collective intelligence of the crowd* [online] Available at:



- <<http://irevolution.files.wordpress.com/2010/05/working-paper1.pdf>> [Accessed 25 April 2011].
- [55] Sheng, V., Provost, F., Ipeirotis, P.G., 2008. *Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers*. In Proc KDD.
- [56] TMCnews,2008,*Moka Chat Launched for Skype Users* [online] (Updated 11 November 2008) Available at:<<http://hosted-voip.tmcnet.com/>> [Accessed 25 April 2011].
- [57]Trados, 2011. *Trados* [online] (Update 26 August 2011) Available at: <<http://www.trados.com/en/>> [Accessed 29 August 2011].
- [58] Transclick INC.,2011.*Transclick Translate, Communicate and Collaborate* [online] Available at: <<http://www.transclick.com/index> > [Accessed 25 April 2011].
- [59] Urriza,2010.*Crowdsourcing Gone Bad* [online] (Updated 16 May 2010) Available at:<<http://urriza.net/blog/2010/05/crowdsourcing-gone-bad/>> [Accessed 25 April 2011].
- [60] UTest, 2011. *Better Apps With Crowdsourcing & The Cloud* [online] Available at: <<http://www.utest.com/webinars/better-apps-crowdsourcing-cloud>> [Accessed 25 April 2011].
- [61] Web Trends, 2011. *A List of Mobile Web Browsers* [online] Available at: <[http://webtrends.about.com/od/mobileweb20/tp/list\\_of\\_mobile\\_web\\_browsers.htm](http://webtrends.about.com/od/mobileweb20/tp/list_of_mobile_web_browsers.htm)> [Accessed 29 August 2011].
- [62] WeLocalize, 2011. *WeLocalize NextGeneration* [online] (Updated 25 August 2011) Available at:<<http://www.welocalize.com>> [Accessed 29 August 2011].
- [63] What is Crowdsourcing, 2011. *The Golden Rules of Crowdsourcing* [online] (Updated 20 January 2011), Available at:<<http://what-is-crowdsourcing.com/the-golden-rules-of-crowdsourcing/>> [Accessed 25 April 2011].
- [64]Wikibhasha,2010. *Wikibhasha beta* [online] (Updated 27 December 2010) Available at:<<http://www.wikibhasha.org/>> [Accessed 29 August 2011].
- [65] Wilson, A., 1983. *Translators on Translating: Inside the Invisible Art*, Vancouver, CCSP Press, 2009.
- [66] Wired,2006. *The Rise of Crowdsourcing* [online] (Updated 1 June 2006) Available at: <<http://www.wired.com/wired/archive/14.06/crowds.html>> [Accessed 25 April 2011].
- [67] World Wide Web Consortunim,2005. *Localization vs. Internationalization* [online] (Updated 5 December 2005), Available at:<<http://www.w3.org/International/questions/qa-i18n>> [Accessed 25 April 2011].
- [68] World Wide Web Consortunim,2011. *Standards* [online] (Updated 20 April 2011), Available at:<<http://www.w3.org>> [Accessed 25 April 2011].

- [69] World Wide Web Consortium,2011. *Notation3 (N3): A readable RDF syntax* [online] (Updated 28 March 2011) Available at:<<http://www.w3.org>> [Accessed 29 August 2011].
- [70] World Wide Web Consortium,2008. *CSS Mobile Profile 2.0* [online] (Updated 10 December 2008) Available at:<<http://W3.org/TR/css-mobile/>> [Accessed 29 August 2011].
- [71] World Wide Web Consortium,2011. *SPARQL Query Language for RDF* [online] (Updated 15 January 2008) Available at:<<http://www.w3.org>> [Accessed 29 August 2011].
- [72] World Wide Web Consortium,2004. *RDF Primer* [online] (Updated 10 February 2004) Available at:<<http://www.w3.org/TR/rdf-primer/#rdfmodel>> [Accessed 29 August 2011].
- [73] ZD Net Asia, 2010. *Mobile Web browsing to outstrip desktop* [online] (Updated 19 October 2010) Available at:<<http://www.zdnetasia.com/mobile-web-browsing-to-outstrip-desktop-62203816.htm>> [Accessed 29 August 2011].