Abstract
This paper describes how the first software localisation venture for the Runyakitara language i.e. the translation of Google interface into Runyakitara was accomplished. Runyakitara is a Bantu language that standardizes four linguistically closely-related languages of western Uganda which are Runyoro, Rukiga, Runyankore and Rutooro. The IT jargon was localized into terminologies that convert the original western material to better reflect African notions. This report describes how all the strings that are inbuilt in the Google interface were translated and how the new terminologies were formulated. It also highlights the major challenges that were encountered and provides some recommendations for future endeavors.

Keywords
Google, Runyakitara, localization, Ugandan languages, software translation, African language
1. Introduction

By 31st December 2009, only one percent of the total number of IP addresses that were issued were from Africa [Number Resource Organisation 2010]. It has also been estimated that less than 10% of the population of Uganda are computer literate [World Fact Book 2008]. This clearly shows that there is a wide digital gap between African countries and the rest. One solution to bridging the digital divide between countries is to address the language barrier that is created when technology is released. People find it harder to embrace any technology if it can only be accessed in a foreign language. A general survey on the World Wide Web quickly reveals how software packages or translation services cater for mainly European and Asian dialects.

The majority of Ugandans are computer illiterate because they have got a phobia for computers and do not easily relate to the IT jargon. Cultural software localisation is one way in which the computer can be demystified so that the semi-literate and non-English speaking community are motivated to embrace and participate in computer technology. Localisation makes it possible for given societies to access software in a language they are most familiar with; their mother tongue. A language is an integral part and one of the most distinguishing feature of culture. It is an indispensable means of communication and the bridge between generations. Cultural localisation acts as a bridge linking modern technology to ancient practices.

Cultural localization has been defined as the process whereby software written in one culture is adapted to the needs and outlooks of another [Keniston 1997]. It involves more than simple translation. The two main components of localisation are translation and modification [Keniston 1999]. Translation is a cross-cultural event involving cross-cultural communication (sharing thoughts) [Bassnett 2001]. It is the linguistic component of localisation and consists of five phases: the initial translation into the target language, back-translation into the original language, comparison between the two versions, adjustments to the version in the target language and incorporation of the now corrected translation into the final localized program [Keniston 1999].

Modification refers to customizing the terminologies in order to fit into the local customs and culture. It involves more “structural” changes, such as scrolling patterns, character sets, box sizes and icons. It can involve cultural aspects and linguistic aspects like dictionary search patterns [Keniston 1999].

Software localization can positively contribute to poverty alleviation by stimulating faster growth of economic activities like e-trading, e-commerce, e-learning, e-banking, email communications, better record keeping etc. Business functionalities
can be enhanced by offering a competitive advantage through increased efficiency, faster information retrieval and improved productivity through better management tools. Software localisation also ensures cultural preservation by encouraging local content development for websites. This is crucial in today’s society that has seen a rapidly increasing western culture influence through modern technologies and globalization.

This work describes the localisation initiative that was carried out as a way to promote the use of ICTs in the Runyakitara speaking communities of Uganda. Runyakitara is a language spoken by four closely related ethnic groups in the western part of Uganda as well as neighboring countries like Tanzania, Rwanda and the Democratic Republic of Congo. Unfortunately the Runyakitara content on the World Wide Web is almost exclusively found only in the “Orumuri” newspaper website [ORUMURI]. We expect to stimulate local content development after the localisation of Google becomes fully operational.

This is the very first Runyakitara software localisation project and has started with the Google interface. This report gives a brief background of the previous translation ventures in Section 2. Section 3 describes the localisation process that was used in order to fulfill that goal and challenges and concluding remarks are given in Sections 4 and 5 respectively.

2. **Background**

2.1 **localisation of Ugandan languages**

The majority of localisation drives world wide have been spearheaded by researchers and universities. African localisation has been scanty and non-reflective of the rich and diverse languages spoken in the continent. The few languages that have been localized are mainly from South Africa plus a bit of Arabic and Kiswahili [Dalvit et al. 2008]. Uganda has seen only two successful localisation projects all coordinated by the Faculty of Computing and IT in collaboration with the Institute of Languages of Makerere University.

The very first attempt to translate software into an indigenous Ugandan language was in 2004 when a team of 7 Ugandans attempted to localize Mozilla into Luganda. They finished the work after one and a half years but before it could be launched, Mozilla upgraded to Mozilla Firefox. Unfortunately, the old version (Mozilla) was not compatible with the new one (Mozilla Firefox) and this meant that a lot of work had to be re-done to ensure compatibility including learning new development /
programming languages. This team was very discouraged and gave up before any CD was released to the public.

In 2008, The Faculty of Computing and IT, Makerere University successfully spearheaded the translation efforts to translate Mozilla Firefox into Luganda [Tushabe 2008]. A 2-day mass translation exercise was conducted in which 320 volunteers participated. This was followed by a 3 month period in which four computer science graduates produced the localized product/software. That very team went ahead and localized the Google interface into Luganda in 2009 [Google Luganda 2009].

Of the 50 languages spoken in Uganda, only Luganda has been localized. This research has localized the Google interface into Runyakitara, another language spoken in Uganda.

2.2. Localisation of google into runyakitara
Google is a powerful website that has been ranked as the most visited website on the Internet [Alexa 2009]. The Google website has got interesting features like a search engine for text, image and video documents, a geographical information system which includes map navigations as well as other amenities to facilitate scholarly research, shopping, email, news and translation services etc. It has emerged as a very powerful tool for social networking, video sharing and advertisement services. Google has an option for multi-lingual access and has already been translated into 130 languages. This has partly been possible due to availability of several resources including funds, online support groups, dedicated teams, technical expertise, exposure and the will to do it.

Unlike Luganda which has approximately three million speakers and is spoken mainly in the central region of Uganda, Runyakitara is spoken in the central and western regions by over 6.5 million speakers spread over nineteen districts of Western Uganda [ETHNOLOGUE]. Runyakitara is a name given to the two major language clusters spoken in Western Uganda, namely, Runyankore-Rukiga, and Runyoro-Rutooro. Bernsten (1998) refers to these languages as four major dialects: Runyankore, Rukiga, Runyoro, and Rutooro. Guthrie (1967) classifies these four dialects as two languages belonging to Narrow Bantu of Niger-Congo family, that is, Nyankore-Kiga (E.13) and Nyoro-Ganda (E.11). Runyakitara is also spoken in the Democratic Republic of Congo (DRC) and some parts of Tanzania under sub-dialects including Ruhaya, Nyambo, Zinza, Kerewe, and Rutuku. All these sub-dialects have more than 50% lexical similarity with Runyakitara. The fact that Runyakitara is also spoken in other countries like in Tanzania, Rwanda and the Democratic Republic of Congo means that its localisation will have a wider impact.
3. The localisation process

The main objective of this research was to enhance social-economic development within the Runyakitara speaking communities by providing an opportunity to conduct business/research using a web browser that is configured in a more familiar language – that is Runyakitara. The specific goal of this work was to translate the Google interface into Runyakitara and to culturally localize the IT terminologies to better reflect into African notions the original western material.

The methodology that was adopted was as follows:

a) Identification of partners.
In order to increase the chances of success and continuity, partners were identified and contacted for possible collaboration. The following organisations participated in this localisation initiative by transferring relevant knowledge and expertise: Makerere University Faculty of Computing and IT, Makerere University Institute of Languages, Google Africa and the Uganda Broadcasting Corporation.

b) Identification of language experts and translators
Software translation is a challenging task because it involves both the humanities and sciences intellectual domains. The services of several professionals were used as a quality assurance measure. Online announcements were made and fifteen of the volunteers who responded were selected. These included five linguistic graduates, five people working in jobs that offer translation services and five ordinary speakers of the language. Ten resource personnel were also invited to participate in the localisation exercise. These included six computer scientists and four lecturers or teachers of the language.

c) Implementation
The strings were distributed among all members of the team for individual translation. A 3-day workshop was then organised in which the team was provided with some basic training in localisation and translation. Background knowledge of computing jargon was explained and intense discussions held to agree upon which words to be adopted, given the complexity and context involved. During the workshop, the team was divided into groups of threes and a group translation conducted. Using a brainstorming approach, all the strings were then peer reviewed by the whole team to come up with the initial translated words. Two linguistic experts were then assigned proofreading tasks to ensure the accuracy and flow of meaning. In some instances, back-translation would be done to ensure that the word/string in a target language gives back the original word/string.
After the linguistic translations were completed, the programmers made the structural changes and integrated the strings into the Google software.

d) Testing
Although the translation and localisation exercise is complete, system testing has not yet been carried out because of delays from Google USA. After it has been availed publicly, a quality and usability study will be conducted in order to identify any errors, anomalies or missing links. This study will take on a quantitative approach in which questionnaires will be prepared by two statisticians who will guide and approve the data collection sub-task. Eight research assistants will participate in the data collection process in which ordinary users in the affected districts will be given a chance to provide their feedback. The questionnaires will then be analyzed and the findings addressed before the final version is publicly released.

4. Results
All in all, four hundred and twenty nine (429) words and two thousand four hundred (2400) strings were localized into Runyakitara. This represented 100% of the total work load. The localized words included:

Names of Languages: Following the principle of transliteration, language names were naturalized to fit into local pronunciation of Runyakitara. Every name of a language in Runyakitara in most cases takes one prefix ‘oru’ as a classifier for language, therefore the translation here was: “Greek = Oruguriika”, “Czech = Oruzech” and “Russian = Orurasha” etc.

Days of the week: Runyakitara has many versions of days of the week. While some call Monday ‘orwokubanza’ others call it ‘ekyokubanza’. Decision was taken based on which one is commonly used in written documents

Company names, product names, patents: As a principle in translation, such names were not translated eg trademarks like Google, Yahoo, Microsoft etc.

Commands like search, open etc and regular IT terminologies previously unavailable were localized as deemed appropriate.

5. Challenges
The major challenges to this localisation drive were the linguistic translations and modifications to suite the local culture. Finding an equivalent term(s) that communicates exactly the meaning in the source language is not a simple task. First of all, Runyakitara itself is a composition of Runyankore, Rukiga, Runyoro and
While the four languages are structurally similar, they have phonological and orthographical differences. For example, Runyankore-Rukiga can translate new as “ekisya” while it is “ekisyaka/e” in Runyoro-Rutooro. Search is “Sherura” in Runyankore but “Ronda” in Rukiga. Adopting a single word from the four different options was really challenging and boiled down to more or less a random choice although generally was biased towards Runyankore-Rukiga.

Secondly, the culture in the computer and information technology language is quite distant from Runyakitara. Runyakitara is a language of full of farming, environment and other cultural related terms. It is quite different from a language of manufacturing and technology. The technical terms and cultural terms of Information technology are not close to Runyakitara language. It was therefore a challenge to coin new words for previously unavailable entities. The solution to some of the technical terms was transliteration. Transliteration is a method of transferring a term as it is from one language and naturalise it into another language. Some terms were transliterated (naturalized) such as “web = weebu”, “Hacker = Omuaaka” and “Tagalog = Orutagaloga”.

Others were localized to synch with local traditions like “Desktop = Aho’orikuhikira “, “Virus = Akahuka” or “Cache = Ekitwero”.

Some words in English can mean two different words in Runyakitara. And translating or localizing a single word sometimes provided insufficient information to make a correct distinction. For example translating the word “Day” could mean either “eizooba” or “orunaku”. Choice of which one should be chosen became challenging since the distinction can only be made within a sentence. Similarly, two or more English terms may have one equivalence in Runyakitara, for example, photo, image, picture can all mean ‘ ekishani’.

Sometimes the team decided to slightly change the original meaning. For example “Search only in = Ronda omu”. This was because although it is more accurate when phrased as “Ronda omu ....... honka“ , a direct translation of “Ronda honka omu” would not make sense, so we opted for a middle ground of “Ronda omu” although this slightly changes the original meaning.

The other technical issue we had to deal with was getting the system to respect the language specific order of the messages prompted or given by the system. This is because in Runyakitara the adjective follows the noun it refers to, while in English the adjective precede the noun. For example “Search LANGUAGE pages” became “Ronda empapura zoru LANGUAGE”
Language change is another phenomenon that we encountered: Runyakitara has greatly been influenced by other languages specifically English, Luganda and Kiswahili. The challenge which such a change comes with, is that one term can be used differently or can have two meanings; one for a foreign language and another for a local language. For example, the term ‘receipt’ is commonly known as ‘risiiti’ as well as ‘akakongi’ in Runyakitara. In other words, some people call it “risiiti”, while others call it “akakongi”. The solution to such terms was to take one that which is commonly used by majority.

Other words were rare and just difficult for the team to translate. These were words like Physics, Astronomy and Planetary Science.

A final comment that we would like to note is that most of the language in Google interface and in the IT field generally consists of command/imperative type of language e.g. search, pick, edit etc. In the Runyakitara culture commands are not polite ways of communication. That is how the language is made. Verbal communication consists of paralinguistic features of tone of voice, intonation, speed of utterance, loudness, patterns of enunciation, and rhythm. We found it daunting to cater for such features in the localisation process.

5. Conclusion
This work describes how the Google interface was localized into Runyakitara. Software localization is a very important component of economic and social development. We expect that a localized Google interface will enable associated communities to interface with the computer easier and break up some negative attitudes and fears associated with using computers in daily chores and businesses. It can stimulate more local content development as well as create bigger opportunities for trade and communication. The localization drive proposed in this project will be extremely beneficial to the majority of indigenous African society especially within the East and Southern African community.

This has been the very first software localisation project in Runyakitara and therefore there is still room for further improvement. Information Technology evolves relatively fast and there is frequent introduction of new words, gadgets and services. These should equally be catered for as well as continual updation of the translated words and strings. Software localisation should also spread to all the languages in Africa, starting with the major ones that are spoken by ethnic groups as opposed to tribes.
Acknowledgements
We would like to thank all the volunteers especially the translators, programmers, and the reviewers for the time and effort that they put in. We appreciate Denis Gikunda and the Google team for the technical expertise. Special thanks go to NUFFIC and The Faculty of Computing and IT, Makerere University for sponsoring this work.

REFERENCES


ETHNOLOGUE, Languages of the World
http://www.ethnologue.com/

GOOGLE
www.google.co.ug


ORUMURI www.orumuri.co.ug

http://cit.mak.ac.ug/staff/tushabe/mozilla.html