

# **Technology preparedness for disseminating flood relief and rehabilitation information to local stakeholders online: Lessons learnt while developing Punjab flood relief website in Urdu**

Huda Sarfraz, Sarmad Hussain, Mahwish Bano, Aniq Dilawari

*Center for Language Engineering, Al-Khawarizmi Institute of Computer Science, University of Engineering and Technology, Lahore*

## **Abstract**

The Punjab flood relief and rehabilitation website <<http://floodrelief.punjab.gov.pk>> has been set up in the wake of the recent floods by the Punjab Provincial Disaster Management Authority (PDMA) in collaboration with the Punjab Information Technology Board (PITB). A variety of information, including financial information, damage estimates, loss statistics and details of relief activities are being updated on a daily basis and being publicized through this website. The website was initially made available in English, and in order to make it accessible on a wider scale, in particular, to the large portion of the population that is not literate in English, it has also been localized into Urdu <<http://floodrelief.punjab.gov.pk/urdu>>. This is a vital measure to ensure that information critical to flood relief and rehabilitation activities is made available in a language medium that is accessible for most of the population. This paper presents the process through which the website is being localized and updated. It will discuss the issues and challenges encountered during the localization, from the aspects of human resources, language, technology and processes, and their solutions. It will investigate how this localization framework can be improved for the future in order to yield the maximum benefit. Finally, some recommendations will be given on the basis of the issues discussed on how to proceed in the future.

## **1 Introduction**

Pakistan has been recently hit by the worst floods in its history. The relief and recovery process is critical in nature, and needs to be made as efficient as possible. Information and communication technology (ICT) can play a key role in facilitating this process. The traditional language for deployment of ICT solutions has been English, but it is evident that in order to

reach the masses, the language medium needs to be one that is understood by the masses. (Sriramesh et al. 2007) presents a survey on the use of ICT for disaster management during recent disasters in the Asian region. Several recommendations are made regarding effective risk communication during disaster situations, and stress is laid especially on the use of local languages given the fact that English is understood by only a small fraction of the population in the area.

The Punjab Provincial Disaster Management Authority (PDMA) in collaboration with the Punjab Information Technology Board (PITB) has taken up such an endeavour by launching a bilingual website to manage relief and rehabilitation efforts in the Punjab area. Frequently updated information regarding damages, relief work, rehabilitation, aid and donations is being made available through the website in both English and Urdu. This information can play a significant role in coordinating the relief and rehabilitation process and in keeping it transparent.

Making flood relief and rehabilitation information available in Urdu in a timely and effective manner ensures that it is available for access by a wider audience, and should significantly increase the impact of the website on relief and rehabilitation operations. This paper therefore focuses on the technological challenges involved in developing and maintaining the Urdu version of the website in particular. By taking the Punjab flood relief and rehabilitation website as a case study, it investigates and makes recommendations in the context of technology readiness in deploying ICT based solutions in local language mediums during disaster situations. For this, the paper first presents a survey of ICT and also localized ICT use in disaster situations. It gives some details on the information being made available through the Punjab flood relief and rehabilitation website. It then goes on to describe the process through which the website was localized into Urdu, and also describes on how website content updates are being managed in the Urdu version. Recommendations are then made based on the lessons learnt through these processes.

## **2 Information and Communication Technology (ICT) in Disaster Management**

This section presents a brief survey of technological initiatives taken to facilitate relief and rehabilitation efforts during recent major disasters in Pakistan and nearby regions.

Sahana is a well known, free and open source disaster management system developed during the 2004 tsunami disaster. It is as a web based collaboration tool, used to organize common coordination tasks in a disaster situation, for example, finding missing people, managing aid, managing volunteers and tracking relief camps (Amin and Goldstein 2008). The system and further information is available from the project website <<http://sahanafoundation.org>>. A deployment is currently also available for the Pakistan floods response <<http://pakistan.sahanafoundation.org/eden/>>.

The Pakistan Space and Upper Atmosphere Research Commission (SUPARCO) used satellite remote sensing and Geographic Information System (GIS) techniques for damage assessment after the 2005 Pakistan earthquake (Ahmed and Bhatti 2006). SUPARCO is also playing a collaborative role in the development of the GIS component of the website being presented in this paper.

Relief and Information Systems for Earthquakes Pakistan (RISE-PAK) is a web based information sharing system designed to facilitate relief efforts in remote areas. It hosted information relating to damages, demographics access and relief, and also served as a portal to feed field information into a central database for better coordination. The system was developed through the collaborative effort of Pakistani and American universities, the World Bank, Pakistan's National Database and Registration Authority (NADRA) and World Online, a Pakistani Internet service provider (World Bank 2007).

Several ICT based initiatives have also been taken to support relief and rehabilitation activities after the 2010 Pakistan floods. Google responded promptly to the flood by setting up a crisis response portal <[www.google.com/intl/en/crisisresponse/pakistan\\_floods.html](http://www.google.com/intl/en/crisisresponse/pakistan_floods.html)> shortly after the initial flooding. The portal includes a collection of resources and tools for flood relief and rehabilitation, including donation details, maps, information on services, and a tool that collects user generated information on missing people. A unique feature of the portal, especially pertinent to the context of this paper, is that it is available in both English and Urdu.

### **3 Punjab Flood Relief and Rehabilitation Website**

This section will give an overview of the content and services available through the Punjab flood relief and rehabilitation website. A preliminary English version was set up shortly after the initial

deluge. This was then improved through a succession of currently ongoing rapid increments. These developments included the addition of new information and deployment of a parallel version in Urdu language. Figure 1 shows a screen capture of the website homepage taken in mid November 2010.

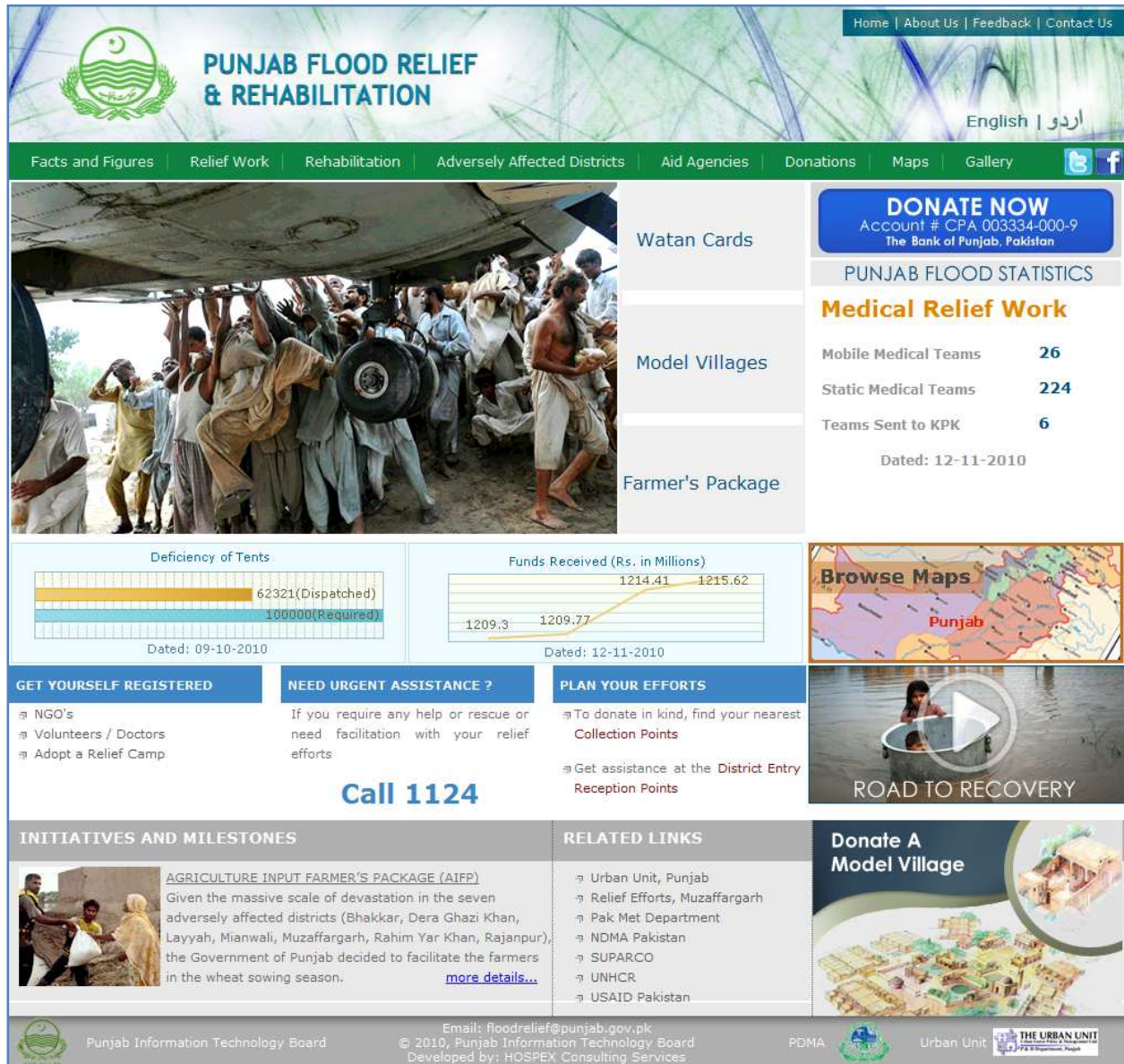


Figure 1. Punjab flood relief and rehabilitation website homepage

The website includes frequently updated and comprehensive statistics on damages, including infrastructure, road network and livestock damages. There is also coverage of relief and rehabilitation processes and statistics. Figure 2 for example shows a screen capture of a

collection points report from district Jhelum taken in mid November. This and many of the other reports on the website are being made available directly from flood affected districts through a portal which enables data entry from the field. The portal is currently facilitating direct updating of reports on dispatch and arrival of relief trucks, inventory of relief goods at collection points, dispatched goods, relief camp details and NGOs working in flood affected areas.

[Back to Main Menu](#)

**Collection Points Report**  
As of: 13-Nov-2010 12:43:35PM

**District : Jhelum**

SNo.	Tehsil	City	Location	Area (Marias)	Is Shed Availabl	Is Warehouse Available	Person Incharge	Mobile No
1		Jhelum	GGHS Nathwala		Yes	Yes	Ambreen Akhtar	0346-9224910
2		Jhelum	G MC Islamia HSS Jhelum		Yes	Yes	Ch. Muhammad Tufail	0544 923007 0333 5823665
3	Dina	Dina	TMA Office, Dina	20	Yes	Yes	Mr.Arif Umar Aziz, DDO(Rev),Dina	0544636550,0333-8013524
4	Dina	Dina	GGES Chak Akka, Tesil Dina, Distt. Jhelum		Yes	Yes	Muhammad Zubair Kiyani AEO (M) Dina	0321-5426148
5	Dina	dina	GGHS BADLOT		Yes	Yes	Safia Sultana	03425286821
6	Dina	Dina	GHS Dina		Yes	Yes	Mian Ashfaq Ahmed	0544 631921
7	Dina	Dina	GGHS PANDORI		Yes	Yes	Shaista Perveen	0544639808 03009573375
8	Dina	Dina	GHS Mota Gharbi		Yes	Yes	Muhammad Anwar	0544 402878 0333 5875255
9	Dina	Dina	GHS Pandori		Yes	Yes	Muhammad Fiaz	0544 639906 0346 5917949
10	Dina	Dina	GHS Ramdial		Yes	Yes	Malik Zafar Mehmood	0331 8803314
11	Dina	Dina	GGHS SAGRI		Yes	Yes	Abida Yasmeen	0544633201 03009563654
12	Dina	Dina	GHS Rohtas		Yes	Yes	Mazhar Hussain Choudhary	0544 685066 0333 5839669
13	Dina	Dina	GGHS SOHAN		Yes	Yes	Naima Yousaf	03435729674
14	Dina	Dina	GGHSS. Dina		Yes	Yes	Nargis Jabeen	0544-630966 0301-5802773
15	Dina	Dina	GHS Saghri		Yes	Yes	Shabir Ahmed	0333 5808492
16	Dina	Dina	GHS Adrana		Yes	Yes	Babar Iqbal	0345 5717506
17	Dina	Dina	GGHS Malot		Yes	Yes	Tehzeem Akhtar	0544-860508 0302-5828487
18	Dina	Dina	Educators Dina		Yes	Yes	Sobia Umair	0333335843846
19	Dina	Dina	Lahore Lyceon Dina.		Yes	Yes	Umar Ayaz Khan	638436
20	Dina	Dina	Raheel Public School Dina.		Yes	Yes	Malik Iftikhar	634666
21	Dina	Dina	GHSS Dina		Yes	Yes	Asad Nadim	0544 630907
22	Dina	Dina	GHS Gatter		Yes	Yes	Sultan pervaiz	0345-5708485
23	Dina	Dina	GGHS. Nakodar		Yes	Yes	Safoora Akhtar	0544-630969 0302-5842779
24	Dina	Jhelum	GGHS KHUKHA		Yes	Yes	Shaheen Kousar	0544445453 03005460146
25	Jhelum	Jhelum	TMA, Office Jhelum	80	Yes	Yes	Mr.Naeem Ullah Bhatti,DDO(Rev), Jhelum	05449270098,03347507755
26	Jhelum	Jhelum	EDO (Education) Jhelum	80	Yes	Yes	Raja Khadim Hussain	0544 9270208, 03016686009

Page 14 of 45

Figure 2. Collection points report for district Jhelum <http://floodrelief.punjab.gov.pk>

The website also has an integrated GIS component which provides a graphical representation of flood affected areas showing the extent of damage to infrastructure. Components of the website are first developed and deployed in English, after which a parallel Urdu version is created. Updates are also first incorporated into the English version of the website, and then into the Urdu version. Section 4 of the paper gives an overview of the process through which the Urdu version of the website was developed. Section 5 then presents the update process for the Urdu website.

## **4 Flood Relief and Rehabilitation Website Localization Process**

The process through which the Urdu version of the website was developed can be termed as localization, which is defined as the process through which a product is customized for a particular language and region. The language in this case was Urdu and the region was the Punjab area of Pakistan, and the product that was being customized was the English flood relief and rehabilitation website. The localization process for the website consisted of the following five basic steps. The first step was identification of the content that needed to be translated into Urdu. The second step was translation of this content into Urdu. The third step was customization of the website layout as per the requirements of the Urdu language. The fourth step was a rigorous quality assurance procedure, after which the Urdu website was deployed, in the fifth and final step. This process requires both technical and language expertise was therefore undertaken by a team consisting of both technical and Urdu language experts working in close collaboration. Further details on the localization process are given in the following subsections.

### **4.1 Content Identification**

During the content identification stage of the initial localization process, the scope was limited to two types of content, textual and image.

All textual content that was available directly on the web pages was selected for translation. Some files that were available from the website were also selected. Files that contained mostly numeric data were not selected for translation. Content that was being made available directly from the field through the portal was not translated. The textual content that was selected therefore included flood relief and rehabilitation reports and other information.

Image content included all those images that needed to be modified for the Urdu website. This included images which either had text embedded in them that needed to be translated, or had the internal layout going from the left to right direction. This direction had to be reversed to go from the right to the left so it adjusted in the layout of the Urdu webpage properly. The website header is an image that can be taken as an example to illustrate both issues. The original English and the transformed Urdu version are shown in Figures 3(a) and 3(b).



Figure 3(a). Website header image from English version



Figure 3(b). Localized website header image from Urdu version

It can be seen that the English text on the header has been translated, and that the image layout has been reversed to conform with the right to left writing direction for Urdu.

Other types of content that could have been translated include for example, videos that are available on the website. At the moment a video available on the website has a voice over in English language. Ideally, this should be translated to Urdu as well, but this has not been covered at this stage of the process. The GIS component of the website has not been included in the localization process at the moment either.

## 4.2 Content Translation

All textual content was translated with the help of OmegaT <[www.omegat.org](http://www.omegat.org)>, a free and open source computer aided translation (CAT) tool. This CAT tool facilitates the translation process by maintaining a memory of previously completed translations and making them available as suggestions for new translation work. It also incorporates terminology glossaries customized for

specific translation tasks, which aid in the efficient production of consistent and high quality translations. Figure 4 shows a translation task in progress using OmegaT.

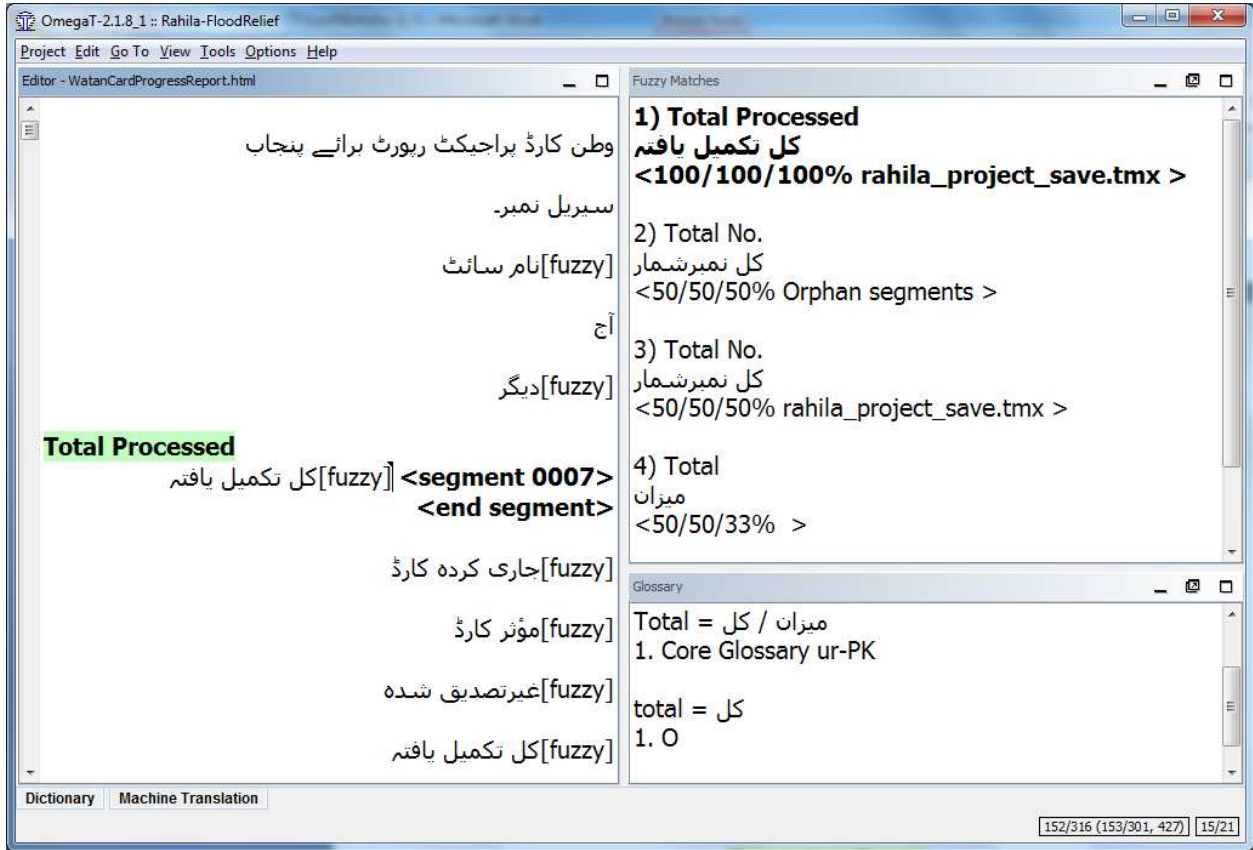


Figure 4. Textual content translation in progress using OmegaT

The main window on the left side shows the sentences that are to be translated. Sentences that are completely translated are being displayed in Urdu. Those that have not been translated but have best match from previous work pre-inserted with a [fuzzy] tag displayed next to them. The sentence that is being worked on by the translator is highlighted. The Fuzzy Matches window on the top right shows other possible matches. The Glossary window at the bottom right shows terminology translations from the glossary compiled for the project.

A basic glossary was compiled at the start of the localization process that included frequently used terminology from the flood and rehabilitation website. As the localization progressed, this glossary was extended to include new terms.



Images, like the one shown in Figures 3(a) and 3(b) were modified using Adobe Photoshop <[www.adobe.com](http://www.adobe.com)>.

### **4.3 Web Site Design Customization**

The design customization process for the Urdu version of the website firstly consists of reversing the layout direction of the web page from the standard left to right direction of English web pages to the right to left direction required by Urdu, because the orientation of Urdu text is from right to left. Secondly, it requires changes in font style and size such that Urdu script is displayed properly.

The layout reversal process includes reversing the basic layout, so that overall layout and basic menu system of the Urdu version goes from the right to the left, and can also include multiple minor details depending on the components of the original webpage. Two examples can be observed in the original and transformed pages shown in Figures 5(a) and 5(b). Figure 5(a) shows a snippet of the page displaying road status. The transformation of this page into Urdu would require the reversal of the general layout, including the menu system, and two additional components as well. Firstly, there is an arrow preceding the title “Status of Roads”, that will be moved to the right hand side as part of the general layout reversal, but in addition, the direction of the arrow will also have to be reversed. Secondly, it should be noted that the flow of information in the table is from the left to the right, with the serial numbers for the data in the left most columns. This direction has to be reversed as well to conform to the direction of Urdu text. Figure 5(b) shows the transformed Urdu version of the website, with the text translated, the general layout reversed, and individual components, arrow direction and table data flow in this case, also catered to.

Home | About Us | Feedback | Contact Us

PUNJAB FLOOD RELIEF & REHABILITATION

English | اردو

Facts and Figures | Relief Work | Rehabilitation | Adversely Affected Districts | Aid Agencies | Donations | Maps | Gallery

→ Status of Roads

**FLOOD EMERGENCY CENTER P&D DIRECTORATE POSITION OF ROADS IN PUNJAB**

SNo.	Name of Road	Flood Affected Road Sections/Kms	Traffic Condition
1.	Jatoi to Karam Dad Qureshi District Muzaffargarh	Km 59-60 (Near Shah Jamal)	Road open for traffic.
2.	Khangarh Shah Jamal Road District Muzaffargarh	Km 6 to Km 16	Road opened for all kind of traffic.
3.	Metalled Road Chotian Sharif to Aman Ullah District Multan	Km 2.50 Road Breached (Near Gardez Pur)	Road open for traffic.
4.	Mettalled Road from Shujabad branch to Aamb Wala District Multan	Km 1.50: Road Breached and water stagnant (Near Suraj Miani)	Road open for traffic.
5.	Mettalled Road from Sher Shah Flood Band to Basti Bhakri Length = 2.75 Km District Multan	Km 2.75, Road Breached and water stagnant (Near Mouza Sher Shah)	Road open for traffic.
6.	Chund Massan to Pir Kot Sadhana road, Length 25.50 KM District Jhang	Km 23 to Km 25.5 Water flowing in breached portion	Road open for traffic.
7.	Adda Rind to Bait Bat Wala Length = 14.35 KM, District D.G. Khan	KM No. 8 Flood breached the road at down stream of Flood bund of Ghazi Chat	Road open for traffic.

Figure 5(a). Status of roads from English website

ابتدائی سطح | جاریہ سطح | رائے دن | رابطہ

English | اردو

اعداد و شمار | ریلیف کام | بحالی | شدید متاثرہ اضلاع | امدادی ایجنسیاں | عطیات | نقشے | گیلری

← سڑکوں کی موجودہ حالت

**سیلاب ہنگامی مرکز پی اینڈ ڈی ڈائریکٹوریٹ، پنجاب میں سڑکوں کی صورتحال**

سڑک کا نام	سیلاب سے متاثرہ سڑک کے حصے/کلومیٹر	ٹریفک کی حالت	سڑکوں کی نمبر
جنونی تا کرم داد قریشی ضلع مظفر گڑھ	59-60 کلومیٹر (نزد شاہ جمال)	ٹریفک کے لیے سڑک کھلی ہے	1.
خان گڑھ شاہ جمال روڈ ضلع مظفر گڑھ	کلومیٹر 6 تا 16	ہر قسم کی ٹریفک کے لیے سڑک کھلی ہے	2.
میٹلڈ سڑک چوٹیان شریف تا امان اللہ ضلع ملتان	2.50 کلومیٹر درائے پٹی سڑک (نزد گردیز پور)	ٹریفک کے لیے سڑک کھلی ہے	3.
میٹلڈ سڑک از شجاع آباد تا آمب والا ضلع ملتان	1.50 کلومیٹر: درائے پٹی سڑک اور کھڑا پانی (نزد سوچ مینانی)	ٹریفک کے لیے سڑک کھلی ہے	4.
میٹلڈ سڑک از شیر شاہ سیلاب بند تا بستی بھکری لمبانی = 2.75 کلومیٹر ضلع ملتان	2.75 کلومیٹر، درائے پٹی سڑک اور کھڑا پانی (نزد موضع شیر شاہ)	ٹریفک کے لیے سڑک کھلی ہے	5.
چنڈ مسن تا پیر کوٹ سدھانہ سڑک، لمبانی 25.50 کلومیٹر ضلع جھنگ	23 کلومیٹر تا 25.5 کلومیٹر	ٹریفک کے لیے سڑک کھلی ہے	6.

Figure 5(b). Status of roads from translated Urdu website

For font style selection, a variety of fonts were available out of which some notable Nastaleeq and Naskh styles were considered. Nastaleeq style is the conventional style of Urdu script used in most Urdu publications. A sample is shown in Figure 6(a). Figure 6(b) shows the Naskh style for Urdu script, which is rarely seen in Urdu publications.

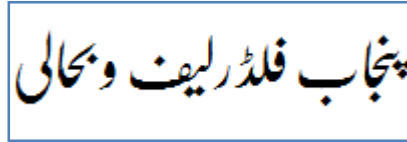


Figure 6(a). Nastaleeq writing style for Urdu

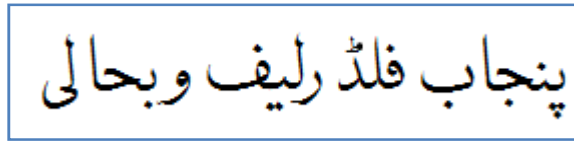


Figure 6(b). Naskh writing style for Urdu

During font selection for the website, the highest priority was given to selecting a font which would display text conveniently and quickly for users. This resulted in the selection of Nafees Web Naskh, available freely from the Center for Research in Urdu Language Processing (CRULP) <[www.crup.org](http://www.crup.org)>. Website users were not expected to have Urdu fonts installed on their browsers, and Nafees Web Naskh could be embedded in the website such that no installation was necessary in order to view it properly. It was also the lightest font identified during the font survey, and its use resulted in web pages that took a minimal amount of time to load. The embedability (Haralambous, 2007) of the Naskh font and some technological constraints in available Nastaleeq fonts make Nafees Web Naskh the best option currently, even though Naskh is not a conventional writing style in Urdu.

#### 4.4 Quality Assurance

The quality assurance procedure consisted of reviewing the website from two aspects. Firstly, the translated text was checked reviewed and secondly the overall website was checked for layout related and other general issues.

Website text was checked for completeness, correctness, understandability and consistency by the Urdu language experts in the team.

For the overall website, it was ensured that links were working correctly. Layout components had to be carefully reviewed to ensure that directionality for all components had been properly set. Layout problems for the Urdu version could be critical in nature because they could result in information becoming unreadable. An example of this can be seen when the bidirectional nature of Urdu text is considered. Although Urdu is written from right to left in general, sometimes it has left to right text inserted within, for example, numbers within text are ordered from the left to the right. This bidirectionality has to be handled from a technological perspective in order to ensure that text is displayed properly and is readable. Figures 7(a) and 7(b) show a possible problem in Urdu text.

Figure 7(a): Urdu text with numbers within the sentence, displayed correctly

Figure 7(b): Urdu text with numbers within the sentence, with display problems

Figure 7(a) shows the text with the correct ordering. Figure 7(b) shows the text displayed incorrectly due to bidirectionality problems. It can be seen at the outset that the text in 7(b) has become left aligned, but if the text is read, errors in the display of numbers can be noted. Specifically the “20” in the date has been misplaced and is causing the information to be misinterpreted. Due to this and other similar issues, the quality assurance process for Urdu websites includes a new dimension which is not applicable in English websites.

#### 4.5 Urdu Website Deployment

The website deployment process consisted of porting the content from a test server after quality assurance to a live server. This was an effective mechanism that ensured accurate delivery of

critical information, and it is recommended that a similar deployment process be used in similar ventures in the future.

## **5 Urdu Website Updates**

The nature of the flood relief and rehabilitation website entails that it needs to be updated frequently in order to be effective. Since its launch, it has been updated with a substantial amount of content on a generally daily basis. The updates are incorporated into the English version on the website first and then the updated English content is passed on to the Urdu website team for incorporation. Other than content updates, because the website is constantly evolving, there are often design updates in the English version that need to be incorporated into the Urdu version as well. The technological aspects of update incorporation are almost exactly similar to the localization process, so will not be looked at in detail in this section. It should be noted though that the translation updates are handled very efficiently using the translation tool described earlier, which retains memory of the already translated text, and only fresh text has to be translated from scratch. Website design changes however usually require repetition of the entire process.

## **6 Urdu Flood Relief and Rehabilitation Website**

The Urdu flood relief and rehabilitation website developed through the process described can be viewed through a link on the English website <[www.floodrelief.punjab.gov.pk](http://www.floodrelief.punjab.gov.pk)> or through a direct URL <[www.floodrelief.punjab.gov.pk/urdu](http://www.floodrelief.punjab.gov.pk/urdu)>. Figure 8 shows a screen capture of the home page taken in mid November.

ایسرانی سفر | بارے متعلق | رائے دی | رابطہ

English | اردو

اعدادوشمار | ریلیف کام | بحالی | شدید متاثرہ اضلاع | امدادی ایجنسیاں | عطیات | نقشے | گیلری

ایچی عطیہ ویس  
 اکاؤنٹ 9-000-000-3334-CPA  
 وبک آف پنجاہ پاکستان

پنجاہ میں سیلاب کے اعدادوشمار

**ریلیف سرگرمیاں**

327	ریلیف کیٹیمپ
9	بہنی کیمپ
155	موٹر بوش
+400	ٹرک
62,321	بہیجے گئے خیمے

تاریخ: 2010-10-09

**وطن کارڈ**

**ماڈل گاؤں**

**ریلیف کیمپ**

**Browse Maps**

Punjab

وصول کردہ غذا (روئے ملین میں)

1047.08
1028.77
1018.7

تاریخ: 14-10-2010

خون کی گھی

62321 (گھنٹہ)
100000

تاریخ: 9-10-2010

**خود کو رجسٹرڈ کرائیں**

این جی او  
 رضاکار / ڈاکٹر  
 ایک ریلیف کیمپ کی کفالت کرنا

**قوری معاونت کی ضرورت؟**

اگر آپ کو اپنی ریلیف کی کوششوں میں کسی قسم کی مدد یا ریسکیو یا سپورٹ درکار ہو تو

**1124 پر کال کریں**

**پنی کوششوں کی منصوبہ بندی کریں**

عطیات دینے کے لیے اپنے قریبی امداد وصولی  
 میگر تلاش کریں

ڈسٹرکٹ افسر، ریسپنشن پلاننگس پر سے معاونت لیں۔

**ماڈل گاؤں عطیہ کریں**

Urban Unit, Punjab  
 Pak Response  
 Pak Met Department  
 NDMA Pakistan  
 SUPARCO  
 UNHCR  
 USAID Pakistan

ریلیف کیمپ میں عید

سیلاب سے متاثرہ علاقوں کے لیے حکومت پنجاب نے عید کے موقع پر کئی تقریبات منعقد کرائی۔ گچہ سادگی کے ساتھ منائی گئی۔ پوئشل ڈسٹریکٹ انتظامیہ اتھارٹی اور ڈسٹرکٹ ایڈمنسٹریشن نے 284,000 کے نزدیک سیلاب سے متاثرہ لوگوں میں کھانا پیش کیا، جبکہ بچوں میں کپڑے اور عید کے تحائف تقسیم کئے گئے۔

ای میل: relief@punjab.gov.pk  
 2010 © پنجاہ انفارمیشن ٹیکنالوجی بورڈ  
 HOSPEM Consulting Services تکمیل کن

پنجاہ انفارمیشن ٹیکنالوجی بورڈ

پنجاہ انفارمیشن ٹیکنالوجی بورڈ

ترجمہ: گچہ ٹیکنیکی وجوہات کی بنا پر اردو ورژن کی تازہ کاری کچھ تاخیر سے دستیاب ہوگی۔

Figure 8. Punjab flood relief and rehabilitation website Urdu homepage

Figure 9 shows another page giving coverage of relief and rehabilitation information for district Dera Ghazi Khan. It can be seen that a wide array of information has been presented using multiple tabs starting at the bottom half of the page. In the screen capture, the NGO section has been opened for viewing.

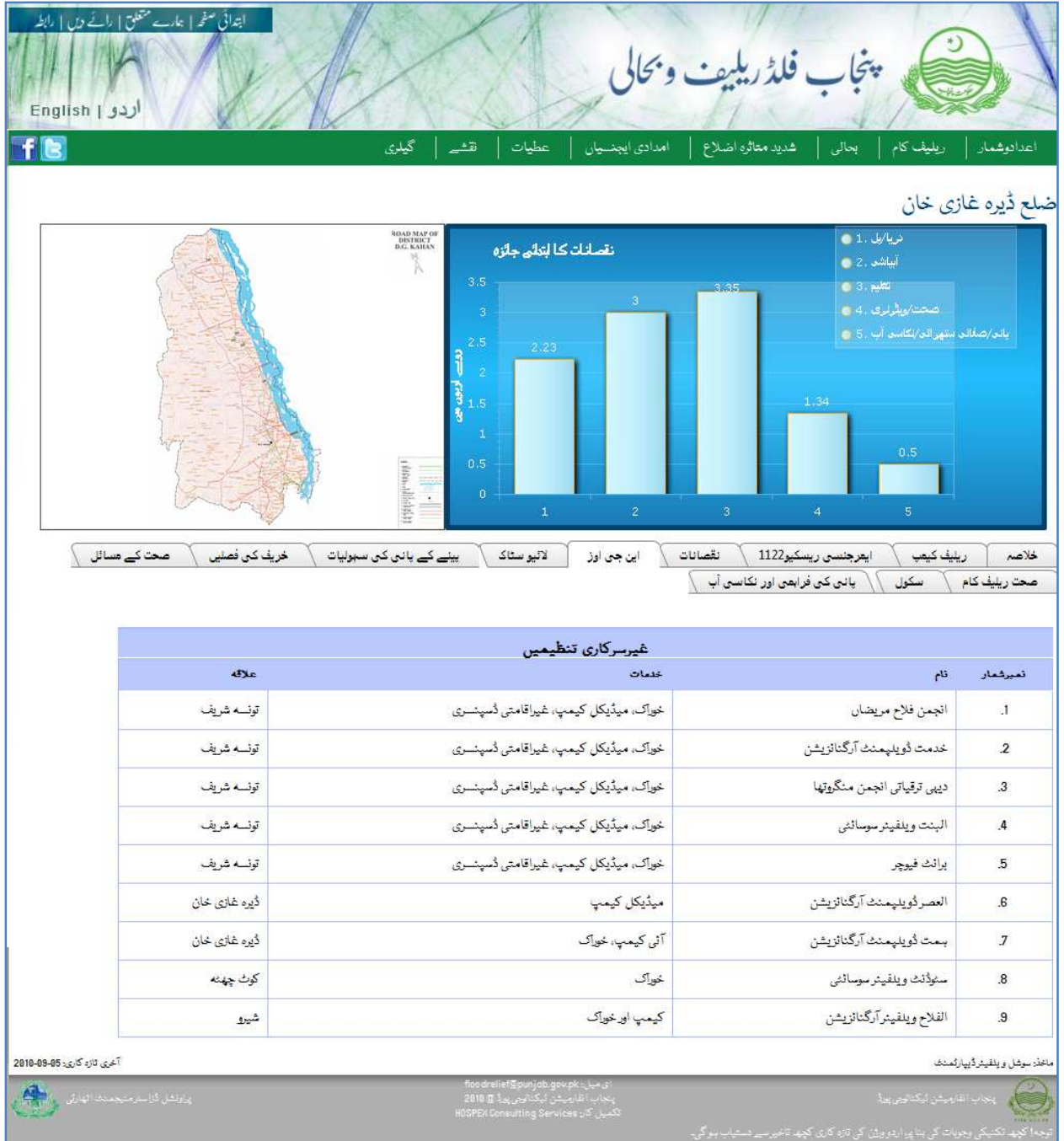


Figure 9. District page for Dera Ghazi Khan

The current localized website contains 7,460 sentences of translated text and several graphics. Apart from the translation memory that has been compiled through the task, a specialized flood disaster management terminology glossary consisting of 462 translated terms has also been

developed. These resources will play a role in efficiently updating the website when more content is added and can also be used for help in translation of other similar content.

## **7 Issues and Challenges**

This section describes two major issues encountered in developing the Urdu version of the website. The first issue is font support for Urdu and pertains to the visual outlook of the website. The second issue covers some translation challenges and pertains to the particular text that needed to be translated for the website.

### **7.1 Urdu Font Support**

As already noted in section 4.3, due to current technological constraints a font style had to be selected which does not conform to the conventional style used in Urdu publications. The Nafees Web Naskh is a practical short term solution, but the ideal situation would be to identify a light and embeddable font Nastaleeq font that can be used in the website. There are plenty of visually appealing Nastaleeq fonts available, but most of them are too heavy for practical use and some have embeddability issues. Any font selected for use within the website needs should be embeddable within the website because Urdu fonts are not expected to be already available on computers being used by website viewers. Users should not be required to install the website font themselves. If the font is embedded directly within the website, website users do not need to explicitly install it and can start viewing website text easily as soon as it loads.

This lack of an appropriate Nastaleeq font is a significant issue because it was noted immediately by users at the launch of the Urdu website. In addition, font usability studies also show that users do not feel comfortable with the Naskh style currently being used in the website, although it is readable. (Urdu Web Interface Guidelines & Conventions Project 2008) reports several users describing the font style as “flat looking”. This shows that the selection of font, although not critical, is still a significant issue and may affect the accessibility of the website.

### **7.2 Translation of Proper Names**

During the text translation process, translation selection for domain specific terms and proper names in particular posed some challenges. Issues related to translation of domain specific terms



were resolved to some extent using several dictionaries and terminology glossaries made available by the National Language Authority (NLA) Pakistan.

No resources were however available to aid in the translation of proper names. The main challenges were posed by place names and government department names. No standard translations were available for place names, and from the multiple place names being used in the website, many were unfamiliar and could not be easily spelled out in Urdu. For example, the place name “Tamewali” is one that is not used frequently, so its pronunciation was not clear to any of the team members. It was not clear what Urdu letter the “T” should be mapped to, because “T” can be used for two letters in the Urdu alphabet, “ت” and “ٹ”. “Tamewali” could therefore be spelled as either “تامے والی” or “ٹامے والی”. There are other English letters and letter combination which can map on to multiple Urdu letter sequences, so this problem was encountered often. The “kh” in “Basti Rind Sakhi” could be mapped on to a “کھ” or a “خ” for example.

In the absence of any standard mapping, the approach taken was to search the Internet for the various options. In many cases only one of the options would yield substantial results which would indicate that it was the correct spelling. This approach did not work in all cases however, because some of the areas being referred to were probably so remote that they could not be found on the web. In some cases, acquaintances were identified who were native to a specific area and were able to provide the proper pronunciations.

Government department names also posed a problem because there was no standard translation guideline. Some were noted to be using full-fledged Urdu translations of their names, while others were simply using their English name transliterated in into Urdu script. This was resolved to some extent through web searches. In case no information was found, the English name was simply transliterated into Urdu script.

These issues can be tackled easily if standard translations of place names and government departments are prepared and released by the government. This can follow a format similar to the list of country names published by the United Nations <<http://unstats.un.org/unsd/geoinfo/9th-UNCSGN-Docs/E-CONF-98-89-Add1.pdf>>.

### **7.3 Update Delay**

One of the challenges faced by the localization team was keeping the Urdu website up to date and parallel with the English website. Due to the nature of the website design and the localization process, in which the Urdu website was updated after updates had been incorporated into the English version, there was always a delay in keeping the English and Urdu websites' content parallel. Although this was reduced to a minimum (roughly about 24 hours), it was impossible to eliminate completely. So the Urdu website content has always been somewhat less up to date than the English website content.

## **8 Recommendations**

This section makes some recommendations based on the work presented in the previous sections. The recommendations are aimed at better equipping ourselves with the tools and technology needed to disseminate critical information to stakeholders in a timely and effective manner. It has already noted that the language medium for such information needs to be local, so recommendations will be made largely in the context of localization.

### **8.1 Language Medium for Information Dissemination**

The first and foremost recommendation, on which the work presented in this paper is based, is to use a local language for dissemination of critical information in disaster situations. This would entail work in localization technology for all languages of Pakistan. This recommendation is also supported in (Sriramesh et al. 2007), which gives an analysis of ICT based disaster management solutions deployed in the Asian region. Also, given the low English literacy levels in Pakistan, it is evident that information disseminated in English will have a smaller audience and therefore a smaller impact. Extending this concept a bit further and taking into account the low literacy level in general it could also be useful to investigate alternate non-textual information dissemination technology, text to speech conversion systems for example. Localization of early warning systems should also be looked into specifically, because they can play a crucial role in mitigating the effects of a disaster.

## **8.2 ICT Awareness and Education**

This paper has investigated the use of ICT in disaster management, but in order for the ICT to be properly used, it is essential to also equip the user group such that they are able to access it. In order to become an ICT-aware society it is essential for general ICT education to be integrated into the curriculum. Information technology (IT) is being taught as an explicit subject at school level, but this is not sufficient. Further steps are needed to ensure that ICT becomes an integral component of education in general, where it is used as a tool for facilitating education in general, as opposed to just a specialized subject. This aspect is also stressed upon in (Shaw and Krishnamurthy 2009). In addition, ICT must also become an integral part of the disaster management curriculum, so that personnel being trained for disaster management are fully equipped to use any ICT based tools available for use during such situations.

## **8.3 Localization Technology Education**

Another aspect of IT education that needs to be looked into is localization technology. Currently, localization technology is not part of any IT or computer science programs available within the country. As a result there are no human resources available who have the proper training to undertake the localization task for any kind of ICT. In order to facilitate the development of localized ICT services, there is a vital need for localization technology to become part of the higher education curriculum. In addition to technical training, there is also a need for specialized training in translation and translation technologies. The website content has been translated by Urdu language experts, but ideally, resources responsible for translation of such a critical nature should have specialized training in translation, where equal attention is paid to both the source and target language for translation, English and Urdu respectively in this case.

Integrating localization technology into the academic curriculum is a long term solution. To address this lack in technology preparedness for the immediate future, specialized training courses should be introduced to cover the technological gap.

## **8.4 Technology Design for Localization**

There is also a need to design solutions with localization in mind. The design for the website presented in this paper was developed without taking the concept of localization into account. The localization process was designed once the basic structure of the English language website

had been created. Due to this, the localization process had to be adapted in a somewhat ad-hoc manner to conform to the English website design. This has resulted in a less optimal process than is ideal, where the Urdu website is updated after the English website has been updated. Due to this, there is an unavoidable delay in updates made to the Urdu website as described in section 7.3. It is therefore recommended that technology is designed with localization in mind from the outset. This recommendation is closely related to the first and third recommendations. If according to the first recommendation, local language ICTs are required by policy, and according to the third recommendation, IT resources are equipped with localization tools and technology as part of their training, future ICT solutions will then be designed with the concept of localization taken into account.

Some work on this is already underway. The PITB is currently working with consultants to assess language localization solutions and to determine localization standards. However, much more work is needed in this direction to ensure adequate technological preparedness in case of future disaster situation.

## **8.5 Content Development Design for Localization**

The content development process for the website presented in this paper was also not optimal for maintaining a bilingual website. The problem was similar to the technical design issue described in the previous recommendation. The process was designed without taking into account the possibility that a parallel version would be required in another language. So, when development on the Urdu version began, ad-hoc measures had to be taken to adapt the existing process. Website content was first created in English for the English language website. This was created using source material that could be in either English or Urdu, depending on the situation. After the content for the English website had been finalized and deployed, it was translated to Urdu and deployed on the Urdu language website. This process could inadvertently cause some Urdu source material to be first translated into English, and then translated back into Urdu by the localization team for deployment in the Urdu version. In order to avoid situations like these, a process needs to be determined where bilingual content is created in parallel, replication of work is avoided, and delays in aligning multiple versions of content are minimized. The communication channels between teams working on various versions also need to be planned out carefully in order to ensure efficient coordination mechanisms. These issues have been resolved

to some extent during the localization of the flood relief and rehabilitation website, but there is still a lot of room for improvement and new solutions need to be investigated.

## **8.6 Budgeting for Localization**

It is strongly recommended to plan in advance for localization whenever information systems are being deployed in disaster situations. This requires determining the languages for which a system needs to be localized and explicitly budgeting for each version, at both the national and provincial levels.

## **8.7 Free and Open Source Software**

The promotion and use of free and open software in future ventures is also recommended as it is a cost-effective option, the effectiveness of which has been proven by notable initiatives in the past. Sahana (Amin and Goldstein 2008), described in section 2 is a prime example. The role of free and open source software in promoting the use of ICT for sustainable human development is explored extensively in (Hoe, 2006). It is also recommended to have personnel trained to deploy and use Sahana at short notice, so it can be put into use immediately in case of future disaster situations.

Another example is the CAT tool OmegaT, which was used in the localization of the flood relief website presented in this paper. It played a key role in producing high quality and consistent translations in an efficient manner, and was available for use at no cost.

## **8.8 Open Source Software Localization**

Sahana and other open source software available for disaster management should be localized for Urdu and other Pakistani languages so they can be used more effectively if needed in the future. Other similar software should be explored and localized, as they have great potential as a feasible and effective solution for disaster management.

## **8.9 Translation Resources Management**

The translation resources compiled through the localization of the flood relief and rehabilitation website, as presented in section 6, can be re-used in case of similar disasters in the future. The translation resources developed include a specialized glossary of translated terms and a translation memory repository. Section 7.2 also suggests for a list of place names and department

names to be released for use in similar situations. It is recommended that these and other similar translation resources should be compiled and then made available publicly so that they can be put into use effectively and in a timely manner in case of a disaster situation.

### **8.10 Technology Medium**

The technology medium explored in this paper is the web. However, research must also be conducted into media that have greater penetration and are also more robust under disaster conditions. One example of this is mobile technology, which has great potential for effective use in disaster situations. (National Research Center U.S. 2007) reports on how mobile infrastructure can be deployed quickly after a disaster situation to re-establish communication links. Mobile based short message services (SMS) specifically can be put into use due to their cost-effectiveness and robustness.

Mobile technology also has great potential in enabling affectees in a disaster situation to directly send in requests and information to a central system. Empowering disaster affectees in this manner will have multiple benefits. However, the use of conventional computers to access such a system during disasters situation is not a viable option. The mobile platform, due to its penetrative power and robustness is an excellent candidate for such purposes. Sahana has a mobile accessibility component that can be explored in this respect. Research should be conducted in this area in order to be better equipped to deal with future possible disasters.

## **9 Conclusion**

This paper focuses on the premise of using local language for dissemination of critical information during disaster situations. Technological readiness to achieve this has been analysed using the development of the Urdu version of the Punjab flood relief website. Finally, some recommendations have been made that will aid in boosting technological capacity to an extent that will enable efficient development and deployment of localized ICT solutions in disaster situations.

## **10 References**

Adobe 2010, Adobe Systems Incorporated, viewed 14 November 2010, <[www.adobe.com](http://www.adobe.com)>.

Ahmed, I. And Bhatti, M.I., 2006, "SUPARCO post earthquake damage assessment through satellite remote sensing and GIS techniques", *Proceedings of International Seminar on Space Technology and Applications*, Islamabad, Pakistan.

Amin, S. and Goldstein, M.P., 2008, *Data against natural disasters: establishing effective systems for relief, recovery and reconstruction*, World Bank Publications.

Center for Research in Urdu Language Processing 2010, Center for Research in Urdu Language Processing, viewed 14 November 2010, <[www.crup.org](http://www.crup.org)>.

Google Crisis Response 2010, Google, viewed 14 November 2010, <[www.google.com/intl/en/crisisresponse/pakistan\\_floods.html](http://www.google.com/intl/en/crisisresponse/pakistan_floods.html)>.

Haralambous, Y., 2007, *Fonts & Encodings*, O'Reilly Media.

Hoe, N. S., 2006, Breaking Barriers, *The Potential of Free and Open Source Software for Sustainable Human Development: A compilation of case studies from across the world*, UNDP and APDIP.

National Research Center (U.S), 2007, *Improving disaster management: the role of IT in mitigation, preparedness, response and recovery*, National Academies Press.

OmegaT 2010, OmegaT, viewed 14 November 2010, <[www.omegat.org](http://www.omegat.org)>.

Punjab Flood Relief and Rehabilitation 2010, Punjab Information Technology Board, viewed 14 November 2010, <<http://floodrelief.punjab.gov.pk/>>.

Sahana Pakistan Floods Response 2010, Sahana Eden, viewed 14 November 2010, <<http://pakistan.sahanafoundation.org/eden/>>.

Sahana Software Foundation 2010, Sahana Software Foundation, viewed 14 November 2010, <<http://sahanafoundation.org/>>.

Shaw, R. and Krishnamurthy, R.R. (Eds.), 2009, *Disaster management: global challenges and local solutions*, Universities Press, India.

Sriramesh, K., Wategama, C. And Abo, F.J., 2007, “The role of ICTs in risk communication in Asia Pacific”, *Digital review of Asia Pacific 2007-2008*, International Development Research Centre.

United Nations Economic and Social Council, 2007, “Country names”, Ninth United Nations conference on the standardization of Geographical names, New York.

Urdu Web Interface Guidelines & Conventions Project, 2008, “Usability Testing Report of Urdu Web Fonts”, University of Management and Technology.

World Bank, 2007, *Building knowledge economies: advanced strategies for development*, World Bank Publications.



## **List of Abbreviations/Acronyms**

PDMA: Provincial Disaster Management Authority

PITB: Punjab Information Technology Board

ICT: Information and Communication Technology

IT: Information Technology

SUPARCO: Pakistan Space and Upper Atmosphere Research Commission

RISE-PAK (fix): Relief and Information Systems for Earthquakes Pakistan

NADRA: National Database and Registration Authority

NLA: National Language Authority

NGO: Non-Governmental Organization

GIS: Geographic Information System

URL: Uniform Resource Locator

CRULP: Center for Research in Urdu Language Processing

CAT: Computer Aided/Assisted Translation

## **Glossary**

**Localization:** The process through which a product is customized as per the needs of a particular language-region pair.

**Font:** A file that defines the display style, shape and size of a letter on a computer based system.

**Font Embedding:** A technology through which a font file can be made part of a webpage so users can view the webpage text as intended without having to install the required font themselves.

**Embeddability:** A property of a font which determines whether it can be embedded into a webpage.

**Computer aided translation tool:** Computer software designed to facilitate the translation process.

**Transliteration:** Conversion of text from one writing system to another.

**Bidirectionality:** A property of writing systems whereby the main text is written from the right to left direction, and some text is written from the left to the right direction.

**Nastaleeq:** A writing style of Urdu script.

**Naskh:** A writing style of Urdu script.

**Nafees Web Naskh:** A font based on the Naskh writing style for Urdu script.

**Translation memory:** A repository of source and target language translation pairs used to facilitate translations using CAT tools.

## **Author Information**

### **Huda Sarfraz**

Email: [huda.sarfraz@kics.edu.pk](mailto:huda.sarfraz@kics.edu.pk)

Huda Sarfraz holds a masters degree in computer science. She has been involved in multiple localization, speech and natural language processing and Urdu localization projects over the last five years.

### **Dr. Sarmad Hussain**

Email: [sarmad@cantab.net](mailto:sarmad@cantab.net)

Dr. Sarmad Hussain, Professor and Head of the Center for Language Engineering, specializes in the area of computer linguistics. He has led a number of projects on language processing, including the PAN Localization Project ([www.pan110n.net](http://www.pan110n.net)).

### **Mahwish Bano**

Email: [mahwish.bano@kics.edu.pk](mailto:mahwish.bano@kics.edu.pk)

Mahwish Bano holds a masters degree in computer science. She has been involved in lexicon development and multiple localization projects over the past five years.

### **Aniqa Dilawari**

Email: [aniqa.dilawari@kics.edu.pk](mailto:aniqa.dilawari@kics.edu.pk)

Ms. Aniqa Dilawari holds a masters degree in computer science. She has been involved in web development projects for the past two years and has recently started working on Urdu localization standards for web development.

## **Postal Address**

Center for Language Engineering, Center for Language Engineering

Al-Khawarizmi Institute of Computer Science

University of Engineering and Technology

GT Road, Lahore