

Internet Governance Forum

Hyderabad, India

"Reaching the Next Billion(s)"

Realizing a Multilingual Internet

3 December 2008

Note: The following is the output of the real-time captioning taken during Third Meeting of the IGF, in Hyderabad, India. Although it is largely accurate, in some cases it may be incomplete or inaccurate due to inaudible passages or transcription errors. It is posted as an aid to understanding the proceedings at the session, but should not be treated as an authoritative record.

>>MARKUS KUMMER: We would like to get started with our proceedings.

The official opening will be in the afternoon, when we go through the formalities. But I think it would be amiss not to say a few words of sympathy to the victims and the families of the victims of the terrorist attacks last week in Mumbai. But we make a formal commemoration in the afternoon at the opening ceremony.

This format has evolved a bit from the two previous formats. We start the morning sessions with panels, and in the afternoon, we have open dialogue. And it is important to us that you react. We have also stepped up our remote participation. All the proceedings in all the rooms will be videocast at all times. We also, for the first time, have live web streaming. So that what you see here in the room, the live transcriptions, they go out on our Web site and remote participants can read what's going on as well. We hope we will get questions from remote participants. And we also hope that we will get questions from the room. And we would also encourage you to make use of the possibility to post your comments and questions on YouTube. We have in the IGF village next door four computers where you can record your messages and comments and questions.

There are changes in the program due to cancellations in the aftermath of the terrorist attacks. And we will communicate them as we go along. Unfortunately, it was not possible to post them on the plasma screens, which are available throughout the conference center this morning. But the staff of the HICC has assured me that they will do their best to update the program as we go along. And I think at the beginning of the next session, I will read out the workshops that have been cancelled today. There are some changes in rooms as well, because the host country reception has been switched from today to tomorrow, some organizers asked, therefore, to change their slots.

But let me now, without further ado, get started. And I would like to pass the microphone to the chair of the session, Mr. Ajit Balakrishnan, the chief executive officer of Rediff.com, which is a company listed on the NASDAQ exchange, New York, a leading Internet company headquartered in Mumbai.

Please, sir, you have the floor.

>>AJIT BALAKRISHNAN: Thank you. You know, good morning to you all. I live in Bombay, so -- and I live right next to the Taj. So all these remarks have great meaning to all of us. My apartment is between the Jewish center and the Taj Mahal hotel, and not far from the Oberoi. So last week was quite a traumatic period and India has changed in a very fundamental way as we will see from the months that go by. In any case, I am not going to be using the PowerPoints, so please put that off.

The theme of today's panel is how to get the next -- the incremental billion Internet users into the world. One of the things which strikes me is if the world is looking to increase Internet users by a billion, India has to contribute at least 250 million of that, I can tell you. And if we don't do that, it is unlikely that you will reach a target of a billion. Now, in -- India's user base right now is roughly 40 million. So it's a mega challenge to take it from where we are now at 40 million to get to 250 million. Now, we got to 40 million after nearly ten years of the Internet. And I can briefly point out in a couple of minutes what some of the obstacles have been so far and it may be something that applies to other countries.

But one level, it is the absence of tools. We do not have enough tools, like parts of speech taggers and so on, and all of you are experts on all of that. We have it in a few languages. Semantic corpuses don't exist, but word net type corpuses exist in Hindi and one or two other languages. Efforts have been made but we're not anywhere near it. But that is not as much as issue in my mind, this will -- these will appear to be the issues which are confronting adoption. But the real issues in India have been when we, for example, have 65 million users of occasional and often, and we have had e-mail operating in 11 languages, but virtually 99% of users prefer to use it in English. In India, there is a set of issues, because practically all of the 300 million young people who aspire to something in this country aspire to learn English. For them, it is partly because the I.T. and I.T. outsourcing boom has created roughly 3 million jobs in the last five years. That's the place people want to be. And if you want to be there, you need to learn English. So English is an aspirational language. Consequently, there is very little interest in accessing the Internet in any other language than that. The additional part is all of Indian higher education, practically all, 99%, is in English. Once you leave school, you have to learn English. Otherwise, you cannot compete in the job marketplace.

The contradiction sheer, however, is practically all of primary and secondary education is in Indian languages. So we are dealing with these complications.

In Indian languages, the issues, of course, have been that we have a large number of them, 16 to 17 major ones. So the market for any of these things is highly fragmented. The single biggest one, Hindi, is probably no more than 30% of the population and not particularly affluent, with the result that early adopters are not very high in Hindi. I think some of these challenges will face the rest. But let us not assume that users want Indian languages. There is no evidence so far the last ten years in the business. If anything, what is happening is what I call the Turkish solution, you know, some years ago it was

decried that in Turkey, you use Latin characters instead of Arabic, and that's always an eye-opener for any of us who visit Turkey from here to see Latin characters. That phenomena is having -- that was decreed. But in India, informally, some of that is well on its way, if you drive through Bombay or Hyderabad, if you see the billboards of Bollywood, which is a famous film industry, practically all the Bollywood films are made in Hindi, all of it. But the billboards and posters are in English. You will see that phenomenon. That's remarkable from the 1980s. It's not anything new. So that is one piece.

The second part is that Nokia, I was checking with the Nokia people, they have experimented a lot with Indian language keyboards and note takers. So, in effect, what we have done is one of our affiliate companies has created a piece of software which does predictive text with the result that you enter Latin characters and what appears on the screen can be said to be Hindi or Tamil or Telugu. There seems to be a kind of underswell movement to using Latin characters to communicate in any of the Indian languages. These are real issues. So I'm sure during today's meeting, you will have an occasion to talk about all of these.

There is a place to do investment in tools. But, to my mind, the issue is well beyond that. I think the consumers don't want it. So I think this is something to think about. Thank you.

>>MIRIAM NISBET: Thank you, Ajit.

Good morning.

My name is Miriam Nisbet. I'm with UNESCO, the United Nations Educational, Scientific, and Cultural Organization.

I'm very pleased to be the moderator for our panel this morning. We have six esteemed panelists who are going to talk about some of the issues associated with the complex issues that our chair has so kindly given us a bit of context for, which is really trying to reach the next billion, and dealing with multiple languages.

A little bit of housekeeping before we start. We are trying something a little different at this Internet Governance Forum this time in terms of the way we're doing our plenary sessions. This, of course, is the very first one. We know that people are still registering, and we know that people will be joining us.

We're going to have until right at about 11:00 to have our discussion this morning on the multilingualism issues. This session will be immediately followed in the same room by another main session on access. And you might be thinking about the connection, the relationship, and how intertwined are the issues of multilingualism and access.

This afternoon, in what's called the open dialogue, which is new, a new format for the IGF, we're going to be talking about both issues, appropriately. And we're really going to, in that session, have the opportunity -- we're not having panelists, as such. It really is an open dialogue and an opportunity for everyone who is participating to ask questions, to comment, to debate.

What we will try and do this morning, we're going to have each of our panelists make a bit of a presentation, a short presentation. We're going to have some questions for them. Hopefully, time will allow for us to have a few questions from the floor. But what I would like to ask all of you to do, as we're going through our presentations and our discussions this morning, please make note of any questions, any comments, any issues that you would like to see discussed this afternoon in the open dialogue. We would really urge you to bring those questions, because we're not sure that we're going to have enough time. Certainly we're not going to have enough time to get into them in the depth to which -- that we hope we will do this afternoon.

So what we're saying, I guess, is, you're a little bit of an experiment. And we are going to be -- have to be a little flexible and spontaneous on how all of this works.

Let me just mention before we go any further, in case we do have people who are connected by remote access, good morning. We welcome you. We're not going to be able to pull you into the Q&A that we have this morning, but you'll have an opportunity also to bring in your questions and comments for the open dialogue.

So, please, for those of you who are listening in remotely, make your comments and questions for that.

We do have interpretation this morning. I have had a little trouble hearing on my headset. But, please, if you would, check your headsets and be ready, because we do have two panelists who are going to be making their presentations in French. Our language interpretation -- and we thank you, interpreters, very much -- we have English, French, Spanish, Arabic, Russian, and Chinese. Those are our six U.N. languages. And we also have Hindi this morning. And we may have another as well. But we at least have those seven.

We think that's quite appropriate for a session on multilingualism.

Let me introduce our panelists. And just to give you an idea of who they are, and then we'll just do a quick framework on what their discussion is going to be, and then we'll go right into those.

Alex Corenthin, who is on my right, is the president of the Internet Society for Senegal. And he's a lecturer at the Polytechnic Institute of the Cheik Anta Diop University of Dakar. And he is particularly going to be talking to us about the challenges of getting African content for the African countries on the Internet.

Manal Ismail is to Alex's right. She is director of the international technical coordination for the government of Egypt. She is a member of the Governmental Advisory Committee to the ICANN board. And she's also a member of what's called the Arabic Script IDN -- Internationalized Domain Name -- Working Group, which has been working for several years on trying to figure out some of the technical and policy problems surrounded with getting Internet domain names in the Arabic script for all the countries that share that. And she'll be talking to us a little bit about some of the challenges and still to come on that.

To Manal's right is Hiroshi Kawamura, who is president of the DAISY Consortium, that's the Digital Accessible Information System. And it's working particularly for people who have print disabilities.

Now, down to my left, and next to Markus Kummer, is Viola Krebs. Viola is founder and executive director of a group, a nonprofit group called IC volunteers. And she's also in charge of the secretariat for MAAYA, that's M-A-A-Y-A, the world network for linguistic diversity.

Next to her is S. Ramakrishnan, who is CEO of C-DAC, which is the center for development of advanced computing here in India.

And he's going to talk to some of -- talk to us about some of the challenges that our chair mentioned in terms of trying to reach the next billion, particularly for those who speak Indian.

By the way, I just wanted to mention, in addition to the languages that we were talking about, in this area alone that we're in in India, there's a language called Telugu, some of you may familiar with that. There are approximately 70 or 72 million people who speak that language. And, yet, trying to ensure that they have even very basic materials in their languages is still something of a challenge. And that's a huge number of people who are not necessarily able to connect easily and use their own language on the Internet.

And then our last panelist, way down at the end -- I can barely see Tulika -- that's why I can't see her. She's not there.

She's here this morning. We met earlier this morning. Tulika will be here, I'm sure, momentarily. Tulika Pandey, she is additional director of the department of information technology for the government of India. And she also will be speaking a bit on some of the issues and challenges related to Internationalized Domain Names. So we'll catch her at some point in here.

So just a little bit of a framework here for our presenters. We're really trying to address sort of three areas associated with multilingualism this morning. One is trying to get content in local languages. As we've just been talking about, language is a necessary vector for communication. And if you cannot access the Internet in your own language or you cannot -- even once you get there, if you do not have content that's available to you, you're going to be stopped right there. Accessibility is -- doesn't mean -- connectivity doesn't mean a lot if you cannot access your own material in language that's relevant to you, to your community, to your country, to your region.

So we're going to talk about some of the issues related with that, with being able to access the Internet, with being able to send and receive e-mail and create content online.

Another issue has to do with localization and availability of tools in order to do that.

And that would include software, it would include the training to use software, it would include hardware. We're talking about being able to make translation of materials localized to meet local needs, and also accessibility issues for people not only with disabilities but also with the language accessibility.

And then the third area that we're going to move into is sort of a new issue that we're not going to get into the technical aspects of it. We really can't. But, rather, the issue of and what are some of the policy reasons behind the push for Internationalized Domain Names, or sometimes referred to, shorthand, as IDNs.

That's really fundamental for accessing Internet sites.

To be able to do that is not something that simply is a matter of turning a switch or just saying okay, from now on, we are going to use -- we're going to use scripts that are different from Latin script.

Rather, these solutions require some very sophisticated technical and technological fixes.

So we want to talk not so much about the technical fixes but, rather, why is this important, why is this a big effort, why are governments involved in that.

So that leads me to just a reminder. We have a lot of workshops. I know that some of them -- some of the workshops have either changed very quickly in the last few days in terms of who the panelists are, in some cases there may be workshops that are canceled. I hope not. But we have a very large number of workshops that get into these various issues in more depth. And we hope that you will look for them.

We have a resource paper on the Web site for this session that lists what the other workshops are. I hope you will take a look at those and look out for those and try and get to those, what we call parallel workshops.

We'll also have resource papers, PowerPoints that anyone has prepared. We are not using PowerPoint this morning, but some of our panelists do have both resource papers and PowerPoints. They will all be available on the Web site for this session.

So, I'm going to turn us over to Alex who will start our discussion, and thank you very much, Alex.

>>ALEX CORENTHIN: .

(no English translation).

>> This is the English channel. We can hear the English channel.

>>MIRIAM NISBET: Thank you, we can. It's sort of going in and out. I don't know if that's a problem with the head set or a problem with where you are, sir.

>> Testing one, two, three, this is the English channel. Can you hear the English channel. Testing, one, two, three.

>>MIRIAM NISBET: Yes, we can.

How about you in the audience? Is there anyone who is unable to hear the English?

>>MIRIAM NISBET: I think we're okay.

>>ALEX CORENTHIN: Very well. May I continue, then? I hope it's working now.

Fine. I will not repeat what I said earlier about the thanks and gratitude.

Let me just try to tell you a little bit about the context of multilingualism in the African framework. And in order to do this, let me help you with a few numbers that I think would be worthwhile pondering on, if we want to remember what our context is.

First of all, I think it's important to recall out of 6,000 languages surveyed, 96% of these are spoken by only 4% of the world population, which I think demonstrates that there's a gap that needs to be filled on the international level since 96% of people are speaking only 4% of the languages, which means that there is a huge number of people who are not involved, as it were, because their language is not being taken into account. One-third, 2,000 spoken in Africa, they fall into four main families of languages, and they are all spoken across several countries, across borders, in other words.

They involve large populations, but rarely do they exceed 40 million per language. Swahili is the exception because we know that the continent has about 800 million population, and it turns out that there are no global languages in Africa aside from these trans-border languages that I just mentioned.

As far as all the other languages go, most of them have few speakers.

We know, for example, that 75% of the remaining languages have not been -- by transcribed, I mean there has been no attempt to transcribe them and they have remained as oral languages.

And for the 25% that have been transcribed, very few of them have been the subject of codification, and the others are still under study or are the subject of transcription that is amateurish.

And this raises problems about the visibility of these languages, both in terms of the hard-copy support, and any support that uses transcription.

Now, a study has shown that out of the languages that have been codified, there is little on the Web about these languages.

Most often, the content is only for a very restricted number of speakers, a community that has a political will and logistics as well to be able to exist.

That is the case for the two most important languages on the continent. Swahili, Hausa and some of the languages spoken in South Africa.

We know that it's always difficult to compare things, but if we look at the visibility of languages on the Net, according to a language observatory project study, we see that do a systemic study on the content on the content on the Net comes from the UK.

(Multiple languages on audio)

Percent, in fact, comes from Africa.

(Multiple languages on audio)

Representative than those two countries, UK and Germany which are not in fact the ones producing the greatest amount of content. But out of the 0.3% from Africa, only 20%, approximately, from South Africa. In other words, 80% of the content from the communities in South Africa and the fact that they are very active on the.

Now, if we look in greater detail in this context.

(lost English translation).

See that one sees much more often the content, symptomatic of the realities of access.

I'm not going to talk too much about access. We'll be coming back to this. Our moderator has already talked about access, and I think that the joint session will make it possible for us to look at the links between these two issues.

4% only of Internet users come from Africa.

I'm thinking of one reason that was raised to explain the limited presence of these languages on the Internet.

But we have to say that the majority of these are using European languages to a great extent.

These are the global communication languages. This is attributable to the fact that they have a need for communication and a need to exist in this way.

Now, if we look at a more detailed way, at the content, from Africa, 0.6% in the common languages, English and French, only 0.6% out of the total in the world comes from Africa.

Now, if we look at you'll these figures, a few questions spring to mind.

Is it a problem of the coding of the characters? I do not think so, because we can use all systems, systems based on pictograms, based on the alphabet, they can all be used on the Internet.

So most of them, languages in Africa, are based on the Latin alphabet. So that issue is one that does not need to be followed too closely. I think we can set it aside.

Number two, is it a question of access. I don't think we need to talk about that because we will be talking about that later.

Is it a question of motivation? Motivation, I would like to focus on motivation, because motivation, to produce content in the local languages is important when the local users can actually use that content. But that motivation can also have a political dimension.

The issue here is that one confuses, often one mixes up national language and official language.

Often, the majority of languages spoken in the African countries, official languages, are languages of the north. The national languages are not truly established as official languages in the educational system of individual countries.

Now, the producers of content are those who have been to the schools, have been involved in the educational system, and they do not have the -- they are not literate in the national language.

So I think that the political level or dimension is that one has to put a strategy, a more vigorous strategy to make sure that the national languages are involved in the educational system so that those who are producing content can take into account the needs, requirements of the illiterate populations in national languages but -- in official languages, but those who are literate in the national languages so content can be produced to be able to preserve the community of languages that only exist if there is a community surrounding a particular language. This low number of users or speakers of a particular language would explain this, I believe.

But the only reason that I think we should stress here is the political will.

Now, my time is up, the moderator is telling me.

I hope I have been at least been able to put on the table an issue that will be of interest to the presenters and to the panel in general.

>>MIRIAM NISBET: Thank you, indeed, and I would like to just mention one of the last points you raised, which is dealing with beyond information literacy and media literacy is getting down and reaching those who are illiterate in the most traditional sense and how you reach them in the first place. And I hope they'll have a chance to talk about some more.

Now I would like to call upon Ramke, Ramakrishnan, to talk to us a little bit about his work here in India.

Thank you, Ramke.

>>S. RAMAKRISHNAN: Thank you, Mr. Ajit Balakrishnan, Miriam. It's a great pleasure.

Good morning, all of you, fellow panelists, and distinguished audience.

Welcome to you all to India.

As you all know, India is perhaps the most diverse country in terms of multilingualism.

We have over 22 official languages, over 2,000 dialects, and four major language families. It's mind boggling in many ways.

And they come from Indo-European, Dravidian, AustralAsiatic and Tibeto-Berman.

22 are just scheduled languages.

What it all means is, further, there is a dichotomy between scriptal languages. Same script, multiple languages, sometimes that otherwise also.

Much work, as Mr. Ajit Balakrishnan was mentioning earlier, since the late 80's, much work has gone into creating corpuses, creating tools, and we still are reaching out to the big numbers, reaching out to the next billion.

We have seen the infrastructure. We have seen the I.T. revolution. And in the last ten years is the (inaudible) revolution.

The key challenge for us is how to -- if we can work on this multilingual challenge, we can reach equal numbers, because there is an energy there.

What is the content issue? Let's look at that.

One is the statistics show, highest content is, today, is in Tamil and Hindi followed closely by Telugu, Urdu, Malayalam, Kannada, and further other languages.

In terms of categories, media dominates. That is newspaper and other things.

Second is in terms of -- what you can call -- how-to kind of categories, then literally, and then some kind of e-governance. Government applications is unique in India.

So given all these challenges, first and foremost there have been emphasis to develop the tools. At the very, very basic level, input, storage, and display. The C-DAC itself and many other players have worked very hard in the last close to two decades in creating these tools.

And we said the 22 meets the process of completion very shortly. And the Unicode has enabled much further acceleration of the progress on these things.

And even respective grammar, dictionary, much progress has happened in these areas as well.

So what it all means is they are much more complex in terms of Web and other topics creation, and a lot more of because of the English as number game was given to you without symmetry, as opposed to 90/10. In terms of 90% of people are in Indian languages, 10 people are English, but then you find the continent is the other way. So we have to swap it, not accept it that that is the way it is, otherwise we will not be able to reach out to the billion people.

You find that good lessons are there in terms of the television, for instance. The regional languages and the movies, for instance, have touched everybody.

So a key question is we need to do so, we can do so.

(scribes lost audio).

Means you have to meet those expectations in terms of tools, in terms of creating content, and it's somewhat nontrivial because of the Indian languages. And we have to address those issues.

You have to work through the process of (inaudible) issues to be addressed.

So those gaps are to be filled. And tomorrow you find in the respect of social network, audio, video, many more, and then the platform also.

So India poses a very, very special problem. You have a billion people. You have a large number of languages. And you need tools. You need to develop contents. And more recently have seen the infrastructure, complete infrastructure opportunities changing. And then the devices come into play. The key challenge is people are saying where are the contents and where are the applications?

Much work has gone into it. Much more work needs to go into it.

These are all Indian language specific.

So therefore, when you talk about multilingualism in Internet, you also know that the revolution has gone into it has great, great potential. Actually, to go into multilingualism in the Internet through mobile. So that is a very great opportunity. We are working on it. And that is a very great need, hunger in there, and we have not even touched the speech aspect of it, the whole thing.

So I think this session is very important from the point of view of there is a percentage of that that are out there, 6 billion people.

We think this is a very important topic. Every country has to grapple with it.

Thank you, Chair.

>>MIRIAM NISBET: Thank you.

Viola.

>>VIOLA KREBS: Excellence, ladies and gentlemen, dear IGF participants. Let me begin by expressing gratitude to he (inaudible) this panel. The host country, of course, my thanks go to the host country and, of course, the Secretariat.

(multiple languages on audio).

The MAAYA network.

(multiple languages on audio).

Be with us today. It's been planned. Adama Samassékou, His Excellency, and Daniel Pimenta.

2008 is the international year.

And this topic, this theme is actually important both for 2008 but also for the future of the Internet.

(multiple languages on audio).

Realize that there were 40,000 spoken languages, languages spoken by humans since.

(multiple languages).

To which, as we have seen, 2,000 are in Africa.

(language other than English on audio).

Truly represented, really represented in cyberspace. Well, ladies and gentlemen, this means that there is work to be done.

(language other than English on audio.)

50 million users compared to 220 million in the United States.

Now, the figures for Africa have already been mentioned. I would now like to focus on five challenges or five, shall we say.

(Language other than English on audio).

In cyberspace go, and I will not be addressing them by order of importance.

The first point I would like to make is the search engines. And also.

(language other than English).

Tools that are available today.

Now, when we realize that the search engines are losing their ability to index a significant portion of content published on the Internet and that publicity is related directly or indirectly to research and can therefore have a significant influence on accessibility to information, then we do, indeed, see that the issue or question of search engines is extremely important for linguistic diversity.

And is not only associated with the order of importance of pages, for example, but also categorization algorithms.

If you look at a search engine such as Google, for example, this is the kind of search engine that uses this system.

But it draws on our own research so as to have links in a targeted fashion.

Now, turning to the translation tools, for the most part, these tools are proprietary. In other words, they are not open systems, open tools.

They are increasingly powerful tools that are available, but there are a number of challenges that still need to be faced in terms of these tools.

The next challenge or point that I would like to raise with you are the Internationalized Domain Name systems, IDNs. And I will be -- I will not be spending a lot of time on IDNs because another speaker will be addressing this issue a little bit later.

The third point or challenge is questions or issues associated with the scripts as well as the hardware that is necessary to be able to use these scripts.

The results of research by (saying name), the languages that are mainly oral languages have to be documented. I think that what we have just seen a moment ago with Mr. Corenthin's presentation illustrates this very well.

UTF8 represents or offers, rather, possibilities to be able to use languages in cyberspace, but the -- but progress, which is relatively slow in this area, does raise a number of challenges.

There are also alternatives to the ICANN system. Net4D would be one example.

The fourth point or challenge I would like to share with you, and this is a very important node, I think, when one talks about linguistic diversity in cyberspace is content, content. How can we give or offer universal access 6789 in other words, an emancipation, as it were, of citizens in the world to be able to have shared knowledge society, a knowledge society, something that is accessible to the greatest number of people.

Perhaps I should touch on standards such as creative commons, and then there's the whole question of royalties and access to information with information commons, and, of course, there is also the fact that we should be thinking a little bit beyond written texts. And especially when we are talking about accessibility to local languages for the most part that are oral languages, not written languages. We have to be able to better integrate tools such as videos, for example, but also sound or audio. And images, pictures, other ways of communicating than the written form.

The last, the fifth and last challenge or point I would like to raise in my presentation is a point that is very closely linked up with what Mr. Corenthin said earlier; namely, the literacy of users numerical or digital literacy.

When we realize that over time there is greater numbers of users who are connected to the Internet but that there are proportionally -- there is proportionally less production in spite of the fact.

When we realize that over time, there is a -- there are greater numbers of users who are connected to the Internet, but that there are proportionally -- there's proportionally less production, in spite of the fact that, today, we have tools such as blogs, B-L-O-G-S, blogs, and that this is very important, because it draws our attention to the user of the Web and makes us aware that it is very important to people to publish content, but to publish content in local languages. And this way, bridges can be built. And the literacy of users will be increased, as well as the publishing of relevant information. That is also something that is crucially important.

Now, in conclusion, I would like to say that I am hopeful. We do see that the Internet today is no longer a Web that is principally used by English speakers. There's a greater diversity of content. There's also a greater diversity of the possibilities that we now have through the Web. And I think that it's safe to say that we can hope that the theme of this conference will be achieved in a future that is -- in the near future. Thank you very much.

>>MIRIAM NISBET: Thank you, Viola. Hiroshi.

>>HIROSHI KAWAMURA: Thank you, madam chair. And good morning, everybody.

The DAISY Consortium is established to meet the requirements of people who are print-disabled, those who are including blind and visually impaired and dyslexic, other cognitive disabilities, and so on.

But the core of the DAISY Consortium's activities is to develop the standard which is open, nonproprietary, interoperable, and free of charge, to be shared worldwide.

At the moment, we are in the division stage of the most current standard to accommodate motion pictures, to include sign language support and to meet the requirements of people who are intellectually disabled and so on.

The DAISY standard is being maintained by the DAISY Consortium, which is the international nonprofit organization legally established in Switzerland.

We are targeting some of the global issues which is critical to be served by DAISY technology, such as textbooks in the classrooms everywhere in the world, including indigenous people's schools. And secondly, the human security concerned information, such as disaster, evacuation training manual. We have a very -- We had a very tragic experience of the tsunami in the Indian ocean which hit many countries in the Indian ocean. And not only those surrounding countries residents, but also many foreign travelers who were killed by the tsunami. So the evacuation manual for each area is crucial, which should be connected, closely connected, with the early warning system. But so far, there is no candidate to solve these issues like DAISY technology.

And HIV AIDS is also very important global issue to be tackled by knowledge-based approach. If everybody knows how to treat the AIDS and HIV, so the current disasters may be minimized.

And the e-environment will be another area of the effectiveness of the DAISY in multilingual context, because DAISY is all about the technology which meets the special requirements of people with disabilities, including all types of disabilities: Physical, that includes visual or auditory; or cognitive, psychiatric. So the requirements of those people are quite unique. That's real-world requirements is a real source of innovation. The synchronization of audio, graphics, and text gives very good flexibility of access to one of the channels, at least. So those who can see, hear, and read text may enjoy everything at the same time. But those who can only touch the Braille may read Braille and share the information at the same time. Those who are dependent on sign languages or symbols may listen to the presentation at the same time, synchronized with text and other media, to join the sharing of knowledge and information.

The current paradigm of media has been neglecting some of the group of people. For video, it's very rich contents. But for those who cannot see the screen, it's almost impossible to understand what's going on. And for audio, for hearing-impaired or deaf people should have captions or sign language interpretation. And for intellectually challenged people, symbols, some of the symbols, are most important to comprehend the contents.

The DAISY may include everything in one standard format. In this way, we are looking at DAISY as the best way to read, the best way to publish, and, thus, create the new paradigm, which will include everybody in society towards the inclusive Internet and inclusive publications.

In closing, I would like to stress that the basic principle of a democratic society, which should be the basis of Internet governance in the future, which is a free and prior informed consent. This is stipulated in the article II of the United Nations convention on the rights of persons with disabilities, but also, it's a very basic human right. And the people who have been excluded from the Internet community or Internet should be included with this principle. And I hope DAISY technology and the DAISY Consortium will be contributing to this end.

Thank you very much.

>>MIRIAM NISBET: Thank you, Hiroshi.

Manal.

>>MANAL ISMAIL: Thank you, Miriam. Good afternoon, everyone.

First, please allow me to apologize to all Arabic-speaking people in the room. Unfortunately, I prepared my notes in English. And with all those terminologies, it would be hard for me to make my points in Arabic and translate all these terminologies and benefit from the translation. So I apologize again. And if you please accept my apologies, and I would be better prepared for any future opportunities.

First, I would like to stress on the importance of the multilingualism theme, especially as we are talking about Internet for all, the theme of the IGF here in India. And when we talk about Internet for all, we definitely don't expect all to be speaking English. So it has to be all languages from all language communities. Especially, we are having the IGF in a country such as India, multilingual and multiscript. So it perfectly suits the overall theme.

And, of course, when we speak about a multilingual Internet, this also has to cover all aspects of the Internet. It's not only IDNs; it's not only the addresses and identifiers; but also it has to do with local content, software, applications, browsers, search engines. So it's, again, a collaborative work for coexisting of all languages, and making sure that -- because, for example, most of the users, when they start browsing, they start from search engines. So it would be more helpful if everything is working in parallel to make sure we end up with a real multilingual Internet, yet a global one.

I participate in two working groups, the Arabic working group on Arabic domain names, and the work group on Arabic script, and would like to share with you our experience from the Arabic community.

The Arabic domain names work group is established under the league of Arab states. And it has to do with looking into IDNs, Arabic domain names, in specific, from a language point of view. We already -- it was very collaborative work. We already defined the Arabic language table.

We already have a list for all the Arabic ccTLDs of the Arab countries. We had a pilot project for testing Arabic domain names dot Arabic domain names, IDN.IDN. The next step was to coordinate with other language communities who use languages that are based on the Arabic script.

And this was a self-organizing work group. It's -- the Arabic script on IDNs work group, ASIWG. And it has been a great experience participating to this specific work group also. It has representation from different stakeholders. We have ccTLD registries, we have gTLD registries, we have government representatives, we have academia, we have technical people. Namely, we have participation from Iran, from Saudi Arabia, from Pakistan, United Arab Emirates, Syria, Malaysia, and Egypt. We have gTLDs, we have, specifically, Afiliat and PIR with us. We have representation from UNESCO, ICANN, and ISOC Africa. And we also have experts participating to the work group, such as Michael Everson and John Klensin, who brings to the group the point of view of the IETF.

Currently, we have representation from language communities, such as Arabic, Persian, URDU, Sindhi, Pashtu, and the -- we are still looking for more participation from other language communities so that we can make sure we are as inclusive as possible.

Our guiding principles in this, we are looking for a solution that can be -- that has to be, of course, standardized, following the IDNA protocol, extendable, to make sure that other languages are smoothly included as they are ready, simple, and transparent from a user point of view, either the registrant or the navigator, fast and easy solution that is not a burden on the registry, and, again, something that works from a ccTLD and a gTLD perspective.

We try to layer our approach in solving the problems. We are working -- we have four layers, basically. It's the protocol layer, the language -- the script layer, the language layer, and on the top is the application layer, which I would say is out of the scope of the work of this group.

We started on the -- the protocol layer, it has to do with rules or things that would be enforced by the protocol. And we were very cautious when. And on the script layer, rules at this level should be agreed to be followed by all registries who would deal with the Arabic script in registering domain names. And the language would be a registry thing. I mean has to do with optional and could differ from one registry to another.

We agreed on the code that we would like to have it as allowed for registration. I mean protocol-valid for registering domain names. And we were very cautious at this layer not to remove anything that might be needed later by other languages.

I won't get into details in this. We have another workshop, I think tomorrow at 4:30, where we can get into more details on how we approached the problems and how we solved it. I just want to stress that we need to cooperate and collaborate in order to coexist. And this might need some compromise and

working together with all the goodwill so that we can achieve this, and to make sure that there is difference between the language and the script. People, they speak languages, they expect to register a domain name in their languages. But on the technical side, it's only script, so we have to be cautious and to work collaboratively to solve all such problems.

Thank you.

>>MIRIAM NISBET: Thank you, Manal.

And now we will turn to Tulika.

Tulika, we talked about you while you were gone, so people know that you're coming in as a representative of the government of India and you're going to talk to us about some of the problems that you have encountered in trying to develop and get into the Internationalized Domain Names.

Thank you.

>>TULIKA PANDEY: Thank you, Miriam. And I apologize for giving you a heart attack by coming in a little late.

I would just like to not take too much time, since everybody has almost covered the issues, and say thank you to the team which developed the Internet Protocol, and thank you to Tim Berners-Lee for exploiting it worldwide. And thank God India has English as one of its associated languages in the official languages, or else maybe we would have missed the entire bus of computers and the Internet.

Just to add to what all my previous speakers have already expressed, and that is to say, to make this Internet inclusive, we are still working in -- step by step, taking one step at a time. So first we spoke about including other languages in addition to English. Then we've started to realize that inclusiveness would mean including people who are not just bound by language, but may have other challenges which may need us to bring in ways and methods to bring them in onto the Internet. That has already been covered by Hiroshi in his talk.

In terms of getting people onto the Internet in their own language, Miriam has given an extensive detail of what we are all going through and what we are trying to attempt to get all of us onto the Internet speaking our language, and yet being able to convey to each other what we imply or what we would like to share in terms of knowledge that we have and the technology that we have developed.

But my last request would be that let us go beyond just talking in terms of getting in the scripts or getting in the languages, because we are still not -- we have still not talked about cultures. We may not be able to include them if we continue to work one step at a time. We are still losing and losing very fast, a lot of tacit knowledge that is there in many of our communities and regions which are getting lost because we have not brought them onto the Internet. They do not find any use of the Internet or the computers, because we have not yet thought of how we would bring them onto the Internet.

It is not a question of what we have to offer, but it is more a question of what would they need from us, the policymakers, the technology developers, and so on.

I will leave this question and close my comment.

Thank you, Miriam.

>>MIRIAM NISBET: Thank you, Tulika.

You leave us with a very hard question, and that is, what do these people, particularly indigenous people, people in the local communities, what is it that they need from us, not what we can give them. Thank you.

We have just a very few minutes. But if we could take even a few questions from the audience, I think it would be a good way to get us thinking about what we're going to talk about in the open dialogue session.

So if anybody would like to come to the microphone, we'd be happy to take -- again, we only have about five minutes. But I'm sure you have been stimulated.

And --

Come on up. Please. And if you would identify yourself. For those of you who are a little reluctant to come up right now, please note your questions, keep your questions, and be ready to come to the open dialogue. All of our panelists and the ones from the next session will be there. And we hope for a lively discussion.

Please, sir.

>>:Thank you, Chair.

Is that all right? Okay.

My name is Louis Pouzin. I am speaking for the EuroInC promoting the use of native languages on the Internet.

>>MIRIAM NISBET: We can't hear you. Could someone be sure that the mike is working.

>>LOUIS POUZIN: I'll start over. Okay. My name is Louis Pouzin. I am speaking for EuroInC, a nonprofit French association promoting the use of native languages in the Internet. Is that all right? Not quite?

Access to the Internet by the next billion is a major issue for this IGF in India. This country is an exceptional instance of a high linguistic diversity, as we heard already from the previous speakers, our Indian friends already have acquired a substantial experience on ways and means to handle the coexistence of such a variety of languages. This experience certainly provides an important contribution

to the goal of the, quote, Internet for everyone, unquote, which remains a slogan unless the Internet is in the language of the users.

Since there exist many users and languages in India, it would be most relevant to hear more, much more, about solutions that are considered to tackle this dilemma and at what level, for example, national, regional, local, such issues are essential for the economic development of countries like India and others at the time of the third IGF, it seems relevant to question its role to foster appropriate vistas. This is part of the evaluation of the IGF capacity to fulfill its mandate. It is, indeed, a subject for the first time in the agenda. So we'd like to hear more discussion on this subject.

Thank you, chairman.

>>MIRIAM NISBET: Thank you. We hope that we can meet that challenge.

I just will note that as -- at the third -- as this -- at the third Internet Governance Forum, the third year in a row, we have raised issues of multilingualism. We're still -- we still have a long way to go. But we thank you, sir, for those comments.

Do -- Okay. Back there.

>>MAWAKI CHANGO: Thank you, Madam Chair, Mrs. Nisbet.

Thanks to all the panel, Mr. Corenthin, Alex Corenthin, (Speaking French) -- of the use of official languages, correction says the speaker, national languages, in the educational system of the individual countries. When you were talking about the problem of multilingualism connected with the Internet.

I wanted to ask you whether you are familiar with the situation in Burkina Faso, for example, where they were fairly advanced in terms of the official languages' use in the educational system and also their policies, because they had a delegate at the ministerial level. Have there been any pilot studies, for example, that could help us, that could teach us something about this issue today?

I'll leave that to the lady who just talked about the Arabic -- I'm sorry, I forgot her name -- the question is how you went about defining the list of country code top-level domains, because often there are issues about deciding whether there's a need for a new convention for defining the country codes by, for example, two letters or three letters or writing the full names, that kind of issue. I would like to hear more about what are the issues that you face by defining your list of country code top-level domains.

Thank you.

My name is Mawaki Chango, from Syracuse University.

>>MANAL ISMAIL: Thank you. Actually, as an Arabic-speaking community, we all agreed on having the Arabic IDN ccTLD strings as the short form of the country names. It was either the two-letter code thing, the full name, or the short name. The full name, of course, is very big to be a top-level domain. And abbreviations, in general, are not used in Arabic language. So we couldn't go for the two-letter code, because even in Arabic, even if it's two letters, sometimes it has a meaning. So it gives another meaning

other than the name of the country. So as an Arabic group, we all agreed on going for the short names. And this is only for the Arabic -- the Arab countries, I mean. We didn't work the whole list. We only did our part of the work.

I hope this answers your question.

>>MIRIAM NISBET: Thank you, Manal.

Let me just remind everyone, we are -- there is another workshop, a parallel workshop, specifically on IDNs, Internationalized Domain Names.

I would ask you, if you are interested, and you want to get into these kinds of questions, that would be a really good place for you to go, because I think we are going to try to address that.

Let me also do this, because I would like to take one more question from the floor, and then our Chair, Ajit Balakrishnan, is going to give us a few concluding remarks, and then we really are going to have to finish this.

But I urge you all, please come to the open dialogue this afternoon. Bring your unanswered questions, and there will be plenty of opportunity to make sure that we get to all of them, and to have a lively discussion.

>>RAM MOHAN: Thank you. My name is Ram Mohan, and I had a question, and perhaps some suggestions for the panel that is assembled here.

First of all, thank you for your comments.

I think one of the things that is missing in this dialogue about multilingualism, multi-culturalism that we want to get on the internet is the need for a framework and a common structure for script and language-based solutions.

We're talking about a problem that begins at the core of the Internet, at the domain name system, and goes all the way to Internet navigation.

There are several things missing. For instance, there is missing a common glossary, a common set of terminologies on how to discuss these issues. Multilingualism is often confused with Internationalized Domain Names. One is not the same as the other.

So what I would urge this panel and other participants in the area of multilingualism is to really think hard about creating a common set of semantics, a common set of terminologies that we can use.

And secondly, urge the creation of some sort of a standard or a shared model for the adoption of scripts and languages online. As Manal was saying, people think languages, but computers work with scripts. And we need to find some common way to bridge that gap. Otherwise, I think we run a real risk of having simply scripts depicted online and not having languages. Some languages may completely miss this transition from an oral world to a digital computer-based world.

Thank you.

>> Yes, thank you.

I will speak French, please prepare your head phone.

National languages in particular for developing countries. From whom we expect more multilingualism than others.

But I would like to know, by way of conclusion, what should we expect from those who are responsible for Internet governance so that we do have a digital universe that reflects cultural diversity? What should be expected from those responsible for Internet governance?

Thank you very much.

>>MIRIAM NISBET: What you should expect I think is that those are the questions we need to answer here and that we are trying to address here during these next several days.

Please participate this afternoon in the open dialogue. Please participate in the workshops over the next several days, and hopefully by the time we get to day four, which is the going forward and the emerging issues, these are some of the things that we might -- we are not going to have all the answers in the next few days, but we can perhaps have a better framework and see where we are going.

May I ask our Chair to bring us to conclusion.

>> AJIT BALAKRISHNAN: You may not like some of the things I am going to say, but nonetheless, when you invite an entrepreneur, you expect original ideas; right?

I have listened with great empathy to speakers who dealt with problems. Some of these are real ones, problems that spoken-only languages have, problems with languages which have written script but too few users have a separate set of problems. Problems which the visually challenged or hearing challenged people have.

All these problems are real, but if you put our feet to the fire and say what is the goal, the goal is to add another billion users to the world, this is not going to help you too much. It will probably add another 20 million or 15 million.

Separately, when we worry about domain names and local language scripts, from a person who comes from the tech side of the Internet, I wonder if we are worrying about problems in the wrong sequence.

It reminds me, a lot of the attempts in the '70s to make X400 a network standard. I don't if any of you are old enough to remember that. In the end, TCP/IP won out. And you must ask yourself why did that happen, because you can sit in a committee and mandate people to do anything you want. People actually do things which are convenient for them.

There is a fundamental reason why TCP/IP is more convenient.

It depends because the kind of original traffic was mostly text based. That's the reason.

Now, using current routers to do what our trade calls level 7 network routing, is not a problem at all if you use URLs for routing. And if you use restful protocols. Since this is going to be translated I am going to spell it out. It is R-E-S-T-ful protocols.

And you have some structure to the rest. It is very easy to route it, because you know that in reality, all URLs are going to be nouns if not pronouns or acronyms. So you finally understand what it is. You find all this energy is being spent on local scripts as URLs for domain names, I encourage you to think about it at all.

Fundamentally, the Internet is not about content. When we sit around in meetings like this, we think that most people read weighty tomes published by the U.N. and others. In fact what they do, mostly young people are on the Internet and what they do is they send messages, brief messages to each other, or post messages on social networking sites or download music or enjoy pictures or video clips. None of this is really particularly language related. Most of the navigational levels have to be in languages. So virtually 90% of the content is text free if you look deeply enough.

And one final thing, if you haven't heard about it, the PC era just ended in the last year and the future of access is mobile. But it's again not going to be text-based mobile. That's another message I want to bring to you from the trenches of the Internet world.

People are frantically working to master the voice-to-text conversion piece, not because we want to convert it for any other reason but for the fact that most of the text, most of the processing software has historically been built around text processing, indexing methods, TDIDF calculations and so on. So the big thing, the big really thing the next five years, whoever gets a breakthrough on a voice-based Internet where you can speak into it and hear things back, that is a big price. If that happens, you will find all your conferences have been wasted.

So I urge you to look at all of this.

I know, I am also a member of several weighty bodies in India, I chair some of them, but I urge you all to read a book by James Scott called "Seeing Like a State." Sometimes a problem looks different when you view it from Delhi or Paris or New York as a state, and the problem looks very different when you see it as a common person.

So my recommendation is please read James Scott "Seeing Like a State" and forget all these efforts. Voice-based Internet is where the future lies. And if national entities have to be pushed to do anything, it's to make sure you make the voice-to-text recognition system accurate. At the moment in India we are not getting results more than 70% accuracy. If you can use the brains and get it to 95%, I think that is fantastic. That will solve all our problems.

Thank you very much.

[Applause]

>>MARKUS KUMMER: This brings us to end of this first panel session.

Those who would like to have asked a question, we can tell them they will have plenty of opportunity it afternoon in the open dialogue.

And the ushers have been going around the room distributing these leaflets where you can write down the question you would like to see addressed and write down your name. We will then try to organize it in the afternoon so we address all the various clusters of issues, and hopefully we will have an interesting discussion.

While the panelists of had panel session are now leaving the panel, I invite the panelists of the next session to join us up here on the podium.

We will arrange the podium. There will be a short, I would call it a technical break, but please don't leave the room. Stay in the room.

We will start as soon as possible after we have changed the panel.

Thank you all, Mr. Chairman, moderator, and panelists.