



Indic-Computing Workshop: Bangalore, Sep 15th-16th, 2002.

Table of Contents

- [Goals](#)
- [Agenda](#)
- [Sponsors](#)
- [Workshop results](#)
- [Downloadable Papers](#)

Goals

The Indic-Computing Group is organizing a two day workshop in Bangalore on September 15-16, 2002. The main purpose of this workshop is build a community of people working in the space of developing local language development tools, applications, and content, to better coordinate their ideas and approaches towards the future of indic-computing.

By building a community of practitioners, developers, linguistic experts and organizers, we hope to share ideas and experiences, facilitate broad discussion about the issues involved and discuss future directions for the field. We also hope that this broad coalition would enable more broad-based and active participation in international standards processes and forums, such as the Unicode Consortium and W3C.

At the end of the meeting, we hope to have assembled a community of technically informed and motivated people to organize and lead the indic-computing development effort into the future. The leadership of this community should be individual driven, technically motivated, and entrenched with youth, vitality and a progressive vision. We hope that you will join us and contribute to taking this effort forward.

Please get in touch with any of the following people if you are interested in participating and contributing to this workshop.

Name	Email Address
Ashish Kotamkar	ashish (at) mithi.com
Tapan Parikh	tap2k (at) yahoo.com
Venky Hariharan	venky1 (at) vsnl.com
Joseph Koshy	jkoshy (at) freebsd.org

-- Ashish Kotamkar (on behalf of the Indic-Computing Team)

Agenda

DAY 1

1. *Experiences of Practitioners* Discuss people's experiences with using local language technologies, particularly highlighting gaps in the technology and the particular reasons certain technology decisions were made. Include NGOs, development organizations, schools, government offices, etc.
2. *Encodings*
Discuss various encoding options, their strengths and weaknesses, and OS and application-level support for local language computing in each. Also discuss the process by which each standard evolves.
3. *Display Technologies*
Discuss various font technologies, their relative merits, and also OS and application-level support for their display and rendering.
4. *Input* *Methods*
Discuss various types of input methods, their advantages, prospective users, and software / driver support.
5. Group dinner and informal discussion.

DAY 2

1. Linguistics
Presentation by groups studying language representation from a linguistic perspective. Discuss the linguistic issues and problems with current encoding systems, fonts and rendering methodologies, and ways these problems can and have been redressed, either by participation in standards amendment procedures, or via other avenues.
2. Tools
Discuss various local language toolkits and APIs, the functionality they provide, and the appropriate ways for those technologies to be included in the future from an application and OS perspective.
3. Organization and Capacity Building
Discuss various ways people have tried to organize this process before, and how we can all work together to consolidate these efforts to build a common platform for discussion and policy.
4. Future Directions
Discuss future research directions for indian-language computing, including speech generation and recognition, machine translation, multi-lingual data retrieval, and other promising research avenues.

Sponsors

The Indic-Computing project wishes to thank:

- [Hewlett-Packard India Software Operation Ltd.](#) for providing the venue and operational support for the workshop.
- Professor Patrick Hall and the [SCALLA project](#) for sponsoring the stay of some of the attendees.



- [Microsoft India](#), for sponsoring dinner for the participants on the 15th.
- [Ekgaon Pvt. Ltd.](#) provided some well-needed late night refreshments.

Workshop Results

by <vijay@ekgaon.com>

A total of 42 participants attended the workshop from all over the country presenting views and experiences from projects on Indian language technologies in the open source and commercial domains.

The workshop had presentations and panel discussions on the following topics:

- Experiences of Practitioners
- Encoding options & standards
- Fonts: Open Type, True Type, Metafont
- Input Methods
- Linguistic issues
- Localisation: Approaches & efforts for Open source

Issues, concerns & Action points

The workshop raised several issues and concerns in taking Indic-computing forward. In the free form discussions among the participants, several key areas on which the future development in Indic-computing depends were identified.

Working groups were formed with participants volunteering to take up active roles in development of these specific areas. These issues, groups and people are listed below:

Short Table of Contents

- [Technology](#)
- [Handbook](#)
- [Technology Map](#)
- [Localization](#)
- [Linguistics](#)
- [Proceedings](#)
- [Administrative](#)
- [Consortium Building](#)

Technology

1. Open type fonts/Unicode standards for Indian languages

Holding OTF training workshop for developing major Indian language OTF fonts

Dr. U B Pavanaja would hold and coordinate the workshop and Mr. Abhas Abhinav would help coordinate for logistics & sponsorship. Efforts would be



made to include participants working in all Indian languages from across the country.

Dr. U B Pavanaja, KGP, Bangalore, <pavanaja (at) vishvakannada.com>
Abhas Abhinav, DeepRoot Linux, Bangalore, < abhas (at) deeproot.co.in>

2. Language Groups

Some of the language groups already formed at the workshop are: Please note (*) indicates Group coordinator, all communication for the particular language should be routed to him/her.

- o Malayalam Group

Rajkumar S, FSF, Trivandrum, <s_raj (at) myrealbox.com> <raj (at) linuxense.com>

Arun M, FSF, Trivandrum, <arun (at) freedevolvers.net>

- o Kannada Group

Dr. U B Pavanaja, Bangalore, <pavanaja (at) vishvakannada.com>
N. Anitha, IISc., Bangalore, <nanitha (at) mgmt.iisc.ernet.in>

- o Hindi & Marathi Group (this development would also affect affiliated Devanagiri-based languages, e.g. Konkani)

G. Karunakar, Netcore, Mumbai, <karunakar@netcore.co.in>
Ravi Kant, Sarai, New Delhi, <ravikant (at) sarai.net>
Ravi Pande, Pune, <panravi (at) yahoo.com>
Prakash Advani, Netcore, Mumbai, <prakash (at) netcore.co.in>

- o Bengali Group

Sayamindu Dasgupta, LUG, Calcutta, <unmadindu (at) soffhome.net>
Ravi Kant, Sarai, New Delhi, <ravikant (at) sarai.net>

3. OTF legal & patent status vis a vis open standards (How open are OTF?)

Mr. Sunil Abraham with Lawrence Liang, an IPR (cyber laws) lawyer would find out and come up with a status report.

Sunil Abraham, Mahiti, Bangalore, <sunil (at) mahiti.org>

4. Standardization & representation at Unicode consortium

Representation in the standardization process for character encodings and in the development of OTF, came up as a major concern amongst the participants. The consensus was that the three issues listed below needed to be tackled at the earliest. Frederick Noronha <fred (at) bytesforall.org> also

suggested that a representative body should first take and concerns of the community with local state governments and the Ministry of Information Technology.

- The framework of KGP, developing Kananda OTF/Unicode standards needs to be studied for this. KGP would make available steps and action taken report.
- Coordination efforts with MIT on Unicode standards. Alerting state governments on various issues in moving towards Unicode. Respective groups for languages would take up these issues with their respective state governments, Group Coordinators please note.
- Interaction on OTF with the international standards consortium/schema & framework to approach for getting standards approved to be developed.

Interim working group:

Venkatesh Hariharan, IndLinux, Mumbai, <venky1 (at) vsnl.com>
Dr. U B Pavanaja, Bangalore, <pavanaja (at) vishvakannada.com>
C.V. Srinatha Sastry, KGP, Bangalore, <sastry (at) ead.cmmacs.ernet.in>
Frederick Noronha, Bytesforall, Goa, <fred (at) bytesforall.org>
Prof. Pat Hall, Open University, London, <p.a.v.hall (at) open.ac.uk>
Sunil Abraham, Mahiti, Bangalore, <sunil (at) mahiti.org>
Ashish Kotamkar, Mithi, Pune, <ashish (at) mithi.com>
Ravi Kant, Sarai, New Delhi, <ravikant (at) sarai.net>
Vijay Pratap Singh Aditya, EkGaon Technologies, Ahmedabad, <vijay (at) ekgaon.com>

On the OpenType font technology front, some of concerns expressed by group included:

- Getting good looking fonts developed
- Development of open source tools for rendering and hinting of OTF fonts (currently OTF development uses proprietary tools)
- Finding font developers for all Indian languages and coordinating their activities

There is no group formed to take up this issue, however it is suggested to the OTF font development groups that they should keep these concerns in mind while finding/employing font developers. However groups are to be formed for various open source tools required for font development. Representation from the community is sought for these. Please contact:

Dr. U B Pavanaja, Bangalore, <pavanaja (at) vishvakannada.com>
Rajkumar S, FSF, Trivendram, <s_raj (at) myrealbox.com> <raj (at) linuxense.com>
Abhas Abhinav, DeepRoot Linux, Bangalore, <abhas (at) deeproot.co.in>

Projects on which working groups have not been formed

1. Input methods



- Keyboards: layout and study for analysis for standards keyboards for Indic languages, short cuts.
 - Voice Input methods
 1. Speech corpora - Resources available with TFIR, Bombay, IIT, Madras
 2. Developing standards for standardization of indian speech corpora
 3. National level action plan for development of speech technology
 4. Hindi text to speech project
<http://sourceforge.net/projects/deepti/>
2. Hyphenation - Word processing & browser support

Handbook

The Indic-Computing handbook intends to contain information useful for developers working at adding support for Indian language processing to computer software. A web platform would allow community to contribute towards building the book. A group is formed under guidance of Mr. Joseph Koshy to take up the task of building of the Book. Mr. Abhas Abhinav and G Nagurjuna have offered to involve students from various Engineering colleges in Bangalore & Mumbai respectively to help build the book, if specific topics are allotted to them.

Joseph Koshy, FreeBSD, <jkoshy (at) freebsd.org>
Tapan Parikh, EkGaon Technologies, Mumbai <tap2k (at) yahoo.com>
Rajkumar S, FSF, Trivendram, <s_raj (at) myrealbox.com> <raj (at) linuxense.com>
G. Karunakar, Netcore, Mumbai, <karunakar (at) netcore.co.in>
Durgesh Rao, NCST, Mumbai, <durgesh (at) ncst.ernet.in>
G. Nagarjuna, HBCSE / FSF, Mumbai, <nagarjun (at) hbcse.tifr.res.in>
Abhas Abhinav, DeepRoot Linux, Bangalore, <abhas (at) deeproot.co.in>

Technology Map

The Technology Map attempts to survey available indian language technologies and provide input to an implementor wanting to implement a real-world indian language solution.

Tapan Parikh, EkGaon Technologies, Mumbai <tap2k (at) yahoo.com>
Joseph Koshy, FreeBSD, <jkoshy (at) freebsd.org>
Rajkumar S, FSF, Trivandrum, <s_raj (at) myrealbox.com> <raj (at) linuxense.com>
G. Karunakar, Netcore, Mumbai, <karunakar (at) netcore.co.in>
Abhas Abhinav, DeepRoot Linux, Bangalore, <abhas (at) deeproot.co.in>

Localization Projects

Major localization projects identified were:

- Mozilla support for Indian languages



Some of the identified problems for the project are listed below. S Rajkumar has taken up the project and would find out the various organizations working on the problem and collate experiences and take the project forward.

1. Translation
2. Local language URLs

Rajkumar S, Trivandrum, <s_raj (at) myrealbox.com> <raj (at) linuxense.com>
Sayamindu Dasgupta, LUG, Calcutta, <unmadindu (at) soffhome.net>

- Localization of ZOPE

Sunil Abraham, Mahiti, Bangalore, <sunil (at) mahiti.org>
G. Nagarjuna, HBCSE / FSF, Mumbai, <nagarjun (at) hbcse.tifr.res.in>

- Open Office localization

1. Recompilation
2. Rendering
3. Translate all the messages in all the different languages

G. Nagarajuna offered to coordinate and allocate resources, Ravi Kant, Durgesh Rao to provide inputs. UI technical dictionary could be made available by Mithi Technologies.

G. Nagarjuna, HBCSE / FSF, Mumbai, <nagarjun (at) hbcse.tifr.res.in>
C.V. Srinatha Sastry, KGP, Bangalore, <sastry (at) ead.cmmacs.ernet.in>
Dr. Hema Murthy, IIT-Madras, Chennai, <hema (at) lantana.tenet.res.in>
Ravi Kant, Sarai, New Delhi, <ravikant (at) sarai.net>
G. Karunakar, Netcore, Mumbai, <karunakar (at) netcore.co.in>
Durgesh Rao, NCST, Mumbai, <durgesh (at) ncst.ernet.in>
Ashish Kotamkar, Mithi, Pune, <ashish (at) mithi.com>

- K Office Localization

Pankaj Kaushal, Sarai, New Delhi, <pankaj (at) sarai.net>

- QT, KDE Localization

C.V. Srinatha Sastry, KGP, Bangalore, <sastry (at) ead.cmmacs.ernet.in>
Pankaj Kaushal, Sarai, New Delhi, <pankaj (at) sarai.net>
Dr. Hema Murthy, IIT-Madras, Chennai, <hema (at) lantana.tenet.res.in>
G. Karunakar, Netcore, Mumbai, <karunakar (at) netcore.co.in>

- GNOME

Arun M, FSF, Trivandrum, <arun (at) freedevlopers.net>
Sayamindu Dasgupta, LUG, Calcutta, <unmadindu (at) soffhome.net>
G. Karunakar, Netcore, Mumbai, <karunakar (at) netcore.co.in>
Tapan Parikh, EkGaon Technologies, Mumbai <tap2k (at) yahoo.com>



- X & FVWM

Joseph Koshy, FreeBSD, <jkoshy (at) freebsd.org> Dr. Hema Murthy, IIT-Madras, Chennai, <hema (at) lantana.tenet.res.in>

- Omega-2

Rajkumar S, FSF, Trivandrum, <s_raj (at) myrealbox.com> <raj (at) linuxense.com>

- Support in FreeBSD for Indian languages

Joseph Koshy, FreeBSD, <jkoshy (at) freebsd.org>

Linguistics & UI Localization

One of the important points that emerged from the workshop was the absence of linguistic groups who could have given some insight on the development of lexical resources and standards for developing UI terminology for various Indian languages. The three important areas (needs) that emerged are listed below. Several participants showed a lot of interest and willingness to work towards the project, however the project is still to be developed further and proper participation of Linguistics groups needs to be taken care of. One of the suggestions for that was for involving Prof. Rajeev Sangal of IIIT, Hyderabad, language group.

- Transliteration Standards should be taken up as an urgent issue - Central Dravidian Institute
- UI standardization for all language
- Lexical resource development for Indian languages

G. Karunakar, Netcore, Mumbai, <karunakar (at) netcore.co.in>
Prakash Advani, Netcore, Mumbai, <prakash (at) netcore.co.in>
Tapan Parikh, EkGaon Technologies, Mumbai <tap2k (at) yahoo.com>
Venkatesh Hariharan, IndLinux, Mumbai, <venky1 (at) vsnl.com>
G. Nagarjuna, HBCSE / FSF, Mumbai, <nagarjun (at) hbcse.tifr.res.in>
Kalika Bali, Picopeta Simputers, Bangalore, <kbali70 (at) yahoo.co.uk>
Manoj R Annadurai, Chennai Kavigal, Chennai, <manoj (at) chennaikavigal.com>
Prof. Rajeev Sangal, IIIT, Hyderabad, <sangal (at) iiit.net>

Proceedings

Post workshop tasks:

1. Subscribing participants to -users and -announce mailing list

Tapan Parikh, EkGaon Technologies, Mumbai <tap2k (at) yahoo.com>

2. Document Proceedings & Outcome



Documentation of workshop proceedings is to be released on Monday, 23rd September 2002. These would also be made available on the web. Efforts would be made to post the proceedings at various mailing lists, Newsgroups to make the community aware of the activity of the group and encourage participation.

Venkatesh Hariharan, IndLinux, Mumbai, <venky1 (at) vsnl.com>
Vijay Pratap Singh Aditya, EkGaon Technologies, Ahmedabad, <vijay (at) ekgaon.com>
Ashish Kotamkar, Mithi, Pune, <ashish (at) mithi.com> Joseph Koshy, Hewlett-Packard, Bangalore, <koshy (at) india.hp.com>

Administration

1. Registering new website www.indic-computing.org & hosting.

Mr. Abhas Abhinav have offered to provide server support and help in building future platform for Indic-Computing.

Joseph Koshy, Hewlett-Packard, Bangalore, <koshy (at) india.hp.com>
Vijay Pratap Singh Aditya, EkGaon Technologies, Ahmedabad, <vijay (at) ekgaon.com>
Abhas Abhinav, DeepRoot Linux, Bangalore, <abhas (at) deeproot.co.in>

2. Operations & Official Administration

Venkatesh Hariharan, IndLinux, Mumbai, <venky1 (at) vsnl.com>
Vijay Pratap Singh Aditya, EkGaon Technologies, Ahmedabad, <vijay (at) ekgaon.com>

3. Communication & Networking

Networking & communication could be effectively handled if the communication channels are clearly identified and distributed. Some suggestions are:

Project Coordination: Joseph Koshy

Spokes Person & Representation to International Consortiums: Tapan Parikh
Industry interface: Ashish Kotamkar User Interface: Vijay Pratap Singh Aditya
Public Relations & Media Interface: Frederick Noronha Liasoning with Government, Linguistic groups & academic institutions: Venky Hariharan

Consortium-Building

The workshop participants felt and Prof. Pat Hall put forward suggestion on formalizing the Indic-Computing into an organizational framework. We intend to take a step in that direction and start building on some larger goal of the community. These are:



- Consortium based approach to represent Indic computing efforts with various stakeholders (linguists, practitioners and end-users - including newspapers, writers, GNU/Linux groups in various parts of the country, firms working on Indian language solutions, Microsoft, etc) in the process. Formalize the organization framework to facilitate the development of Indic computing
- Representation of Indic developers as a body with Mozilla, KDE, Gnome etc platform and get the patches to support Indian language included in the support. Network with international efforts (e.g. i18n) attempts to find Indian language solutions to GNU/Linux and other Free Software.
- Coordination and closure interaction with Linguistics & NLP groups in Indian languages, LLI (Linux Localisation Initiative). LUGs in different parts of the country
- Ireland localization industry is the biggest international localization industry. Indic computing should represent and understand the model of Ireland localization industry - Prof. Pat Hall agrees to fund visit for two participants of Indic Computing
- Locate others working on this field, whom we may not be aware of.
- Build bridges with expatriate communities involved in developing local language solutions or looking towards contributing towards the initiative

The team that would take up formation of the consortia and develop the framework is as under. All those members instead could get in touch to provide help in building this effort further.

Joseph Koshy, Hewlett-Packard, Bangalore, <koshy (at) india.hp.com>

Tapan Parikh, EkGaon Technologies, Mumbai <tap2k (at) yahoo.com>

Venkatesh Hariharan, IndLinux, Mumbai, <venky1 (at) vsnl.com>

Abhas Abhinav, DeepRoot Linux, Bangalore, <abhas (at) deeproot.co.in>

Ashish Kotamkar, Mithi, Pune, <ashish (at) mithi.com>

Vijay Pratap Singh Aditya, EkGaon Technologies, Ahmedabad, <vijay (at) ekgaon.com>

Frederick Noronha, Bytesforall, Goa, <fred (at) bytesforall.org>