

Centre of Excellence in Marathi Language Computing: A Case Study

1. Abstract

DIT, Government of Maharashtra and C-DAC, GIST Group Pune signed the MoU for Centre of Excellence in Marathi Language Computing on 18th May 2012. The purpose of this Centre is to assist Government of Maharashtra and its various consulting bodies on various issues related to Marathi Language computing. Suggest them the best practices to conceptualize and develop state of the art web and desktop applications. Train Government employees for effective usage of desktop and web application and in doing so create content which is error free and stick to the latest international Unicode standards. Data thus created is portable and usable to the entire world. CoE team worked in close tandem with DIT to get maximum results in short time intervals.

As part of this unique, first of its kind, initiative in India, reading material and links of important sites are also provided on above website, which would facilitate learning of basic and fundamental concepts of language technology. Apart from the reading material available on this page website, links for additional reading material are mentioned along with various free software's and fonts to download. The project website is <http://coe.maharashtra.gov.in>

2. Key words and tags

STANDARDIZATION
 LOCALIZATION
 TRAINING
 RESOURCE ENRICHMENT & DISSEMINATION
 TESTING & BENCHMARKING
 CERTIFICATIONS & AUDIT

3. Note to Practitioners/Instructors

This case study emphasizes on the need to follow local and international standards to standardize Marathi language on digital platform. Why?

- To reach to the masses there is a need for local language
- Language plays a vital role in successful implementation of e-Governance and targeting rural population effectually
- User friendly interfaces are pre-requisite for successful implementation and delivery of any e-Governance initiative

Marathi is the State Official language in Maharashtra thus; the e-Governance initiatives need Marathi language support. Keeping this dictate in mind, Marathi is essential in all e-Governance initiatives. DIT, Government of Maharashtra has been working relentlessly towards implementation and standardization of the initiatives in Marathi language. Marathi needs to co-exist with English complying with national and international standards for information interchange. Wherever local administration finds it necessary

and expedient to introduce other regional languages, they may do so in a phased manner complying with all appropriate established standards.

4. Project Context

The project was conceived keep the core challenges in mind when implementing language support on digital medium. A few of them are as follows

- Unicode - Ignorance of information about Unicode led to formation of non-standard data.
- GIGW, W3C, Localization standards not followed while designing application.
- Non Standard Inputting mechanism- People well versed with Typewriter keyboard and is reluctant to change to INSCRIPT Keyboard (BIS standard)
- Legacy Data - Legacy Data in proprietary font format not visible in latest operating systems across state.
- Geographic coverage was vast and the projects main aim was to encourage use of ICT even to villagers who find it difficult to cope with the English terms used in various government initiatives.

5. Project Overview

Considering the expertise of C-DAC GIST in the areas of Indian language computing, we wish to propose CENTRE OF EXCELLENCE IN MARATHI LANGUAGE COMPUTING. This center will be set-up by DIT, Government of Maharashtra in technical collaboration with GIST Group of C-DAC.

Objectives of CENTRE OF EXCELLENCE IN MARATHI LANGUAGE COMPUTING
The center will ensure compliance of the following mission critical objectives at state level:

STANDARDIZATION

- Standardization of fonts, inputting, storage and display mechanism
- Usage of standardized Unicode compliant tools for Marathi in all e-Governance applications
- Implementation of e-Governance standards adhering to web guidelines proposed by NIC
- Evolve best practices for e-Governance application development
- To facilitate DIT, Government of Maharashtra for active participation in W3C, Unicode consortia
- Compliance with W3C standards w.r.t websites and web applications
- Support to ensure interoperability of all the applications across platforms and multi modal devices

TRAINING

- To train the developer community on the standards (Internationalization and Localization) that should be adopted while developing application having Marathi support
- Train the stakeholders

- Advice to DIT approved agencies/users associated with the use of Marathi language products

RESOURCE ENRICHMENT & DISSEMINATION

- Best Practices/ guidelines for creating knowledge repository of various resources available in Marathi Language (e-books, audio books, interactive books, etc)
- To network with other Resource Centers working in areas related to Marathi language enrichment, potential user agencies, and other educational institutes
- Development of Marathi lexicons, standardized dictionaries, Thesauri, FUEL terminologies banks, CLDR
- Design and Development of Localization projects management framework.
- Help localize empanelled open source software integrated with standard Government of Maharashtra templates
- Study existing applications and evolve best methodologies for migrating to new environments keeping in mind the standards established
- Organizing periodic IT localization clinics

TESTING & BENCHMARKING

- A test lab for testing the quality of the products/services/applications will be set up
- Norms and standards laid down for the products/services/applications will be tested in the lab and thorough evaluation of the product will be carried out.
- Testing could be carried out in co-operation with standard bodies

CERTIFICATIONS & AUDIT

- Certification of the all applications/web portals developed after confirming to the testing & benchmarking standards
- Audit the applications/web portals developed with respect to the standards established
- Certification of trained localization experts, localization developers & users

BENEFITS : POST IMPEMENTATION SCENARIO

- Unicode - Continuous training and knowledge dissemination helped to clear the obvious advantages of creating contents and application based on Unicode standards. Data exchange and portability in Marathi becomes easy. 15000 people trained till date.
- GIGW, W3C, Localization standards - Vendors were trained on these standards which helped in considerably improving the quality of the applications developed. Around 60 important websites belonging to Government of Maharashtra are adhering to these national and international standards and the numbers are increasing every day as it's a continuous process.
- INSCRIPT Keyboard- INSCRIPT Typing tutor was made available for learning and also for free download on CoE website. Competition and trainings were organized to promote the INSCRIPT standard. Workshops organized with various MNC's like Apple, Microsoft, IBM, Nokia, Samsung helped in clearing their doubts about implementing language technology in their applications and devices.
- Legacy Data - Custom converters in office automation and offline applications prepared which solved the problem to a great extent. Now data exchange happens in Marathi Unicode and can be shared across various applications without fearing any loss.

6. Issues and Challenges faced during Implementation

- Legacy Marathi data based on various proprietary standards not compatible with latest OS.
- Low awareness amongst users and developers on the tools and technologies available for Unicode content creation.
- Typewriter method of inputting creating nonstandard content.
- Resistance of the stakeholders to adapt to various national and international standards which are used worldwide.
- Various agencies empaneled following nonstandard practices for developing websites and applications.
- No support of Marathi Unicode in DTP applications.
- Penetration of standards till zilla and taluka level difficult due to shortage of skilled trainers.

Solutions to the above challenges were overcome as follows

- Unicode - Ignorance of information about Unicode led to non-standard data. Continuous training and knowledge dissemination helped to clear the advantages
- GIGW,W3C, Localization standards - Vendors were trained on these standards which helped in considerably improving the quality of the applications developed
- INSCRIPT Keyboard- People well versed with Typewriter keyboard and are reluctant to change. INSCRIPT Typing tutor was made for learning. Competition and trainings were organized to promote the INSCRIPT
- Legacy Data - Legacy Data in proprietary format not visible in latest operating systems across state. Custom converters in office automation and offline applications prepared which solved the problem to a great extent
- Alignment of initial plans of the Centre with the IT implementation policy was done with the help of DIT team to increase throughput
- Penetration – Using the V-SAT Channel in Yashada, training was imparted till zilla and taluka level. This was the quickest and easiest way to reach masses.

7. Key Lessons

- A push from State's Information Technology department helps in implementing standards to a large extent.
- Continuous auditing and training is compulsory to maintain the standards for data and applications created.
- GAP Analysis is to be done compulsory at the start of any project
- Tools and Technologies have to be customized based on the present user requirements
- Satellite / Video conferencing equipment till zilla / panchayat level helps in promoting and proliferating knowledge in the shortest time span.
- Aligning strategies as per the states e-governance plan helps to manage the project more efficiently.
- Free and Open source tools available on websites helps in covering larger demography.

- Training cybercafé owners on free and open source tools in Marathi created a business opportunity which helped to increase the awareness.
- The project may face opposition and might be slow on impact initially, but once the standards and methodologies are accepted and implemented, the impact rises exponentially.

8. Methodology adopted for Case Writing

This case study is prepared by Mr. Swapneel Raut – Project Leader, Centre of Excellence in Marathi Language Computing in consultation with Mr. Mahesh Kulkarni – Associate Director and HoD, GIST Labs, C-DAC, Pune. The details are collected from all the stakeholders like trainers, developers, designers, employees of state government, users.

9. Case Fact Sheet

C-DAC, GIST along with DIT, Government of Maharashtra is the joint owners of the project. This project as stated earlier is for the benefit of all the users who want to use Marathi language in digital medium. The deliverables are based on the requirements of Government of Maharashtra and articulated by Secretary IT. The requirements are broadly classified in the form of

- Training – Developer and User
- Tools and Technologies required for localization
- Standardization
- Certification and Audit
- Promotion and Proliferation

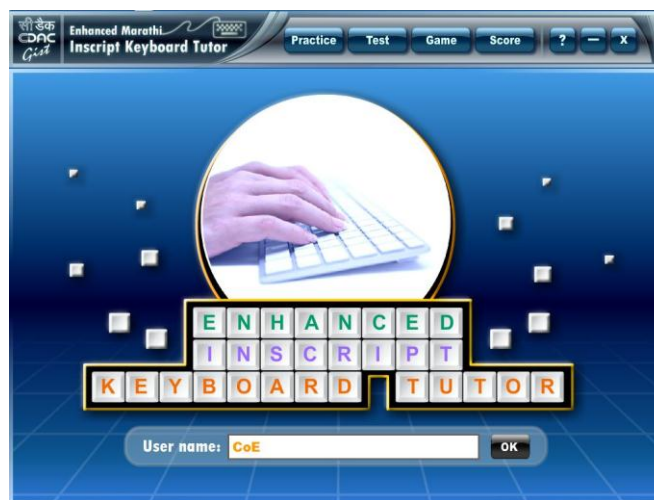
The Centre delivers in all the above areas and also suggests best practices for DIT so that it can reshape its IT E-Governance policy accordingly.

Screen Shots of some tools deployed in various applications and portals in government of Maharashtra.

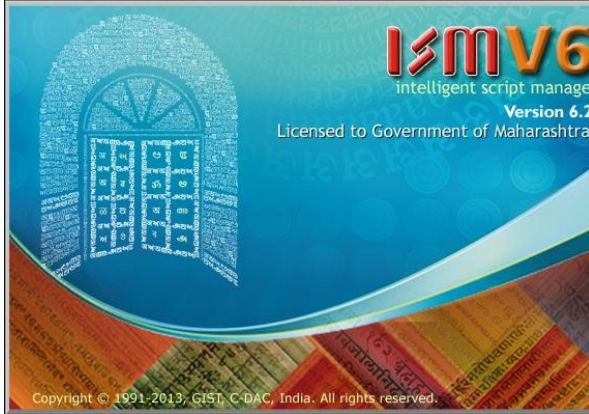
CoE Website



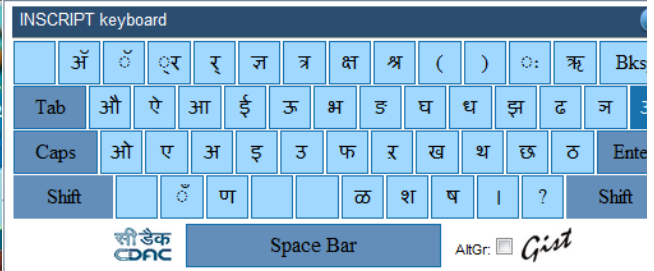
Typing Tutor



ISM 6.2

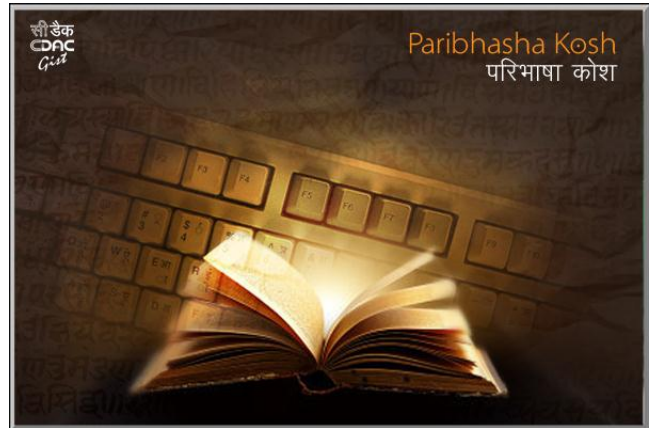


Floating Keyboard – INSCRIPT



Sakal Marathi Font in open source

Paribhasha Kosh



Authors

Institute/Organisation	Centre for Development of Advanced Computing and Directorate of Information Technology, Government of Maharashtra
Address	Centre for Development of Advanced Computing, GIST Group, 4th Floor, Westend Centre III, Aundh, Pune - 411007 Directorate of Information Technology, Government of Maharashtra, Room 514, Fifth Floor, Annexe Building, Mantralaya, Mumbai.
Project Author	Mr. Mahesh Kulkarni Associate Director and HoD, GIST Group, C-DAC mdk@cdac.in
Project Co-Author	Mr. Rajesh Aggarwal Secretary IT, Government of Maharashtra Sec.it@maharashtra.gov.in
Project Commencement Date	18th May 2012
Project Completion Date	17th May 2015