



Microsoft® Research

FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS



Microsoft® Research

FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

**Addressing Societal Challenges Through Innovation
and Partnerships – Microsoft Research in India**

P. Anandan
Distinguished Scientist and Managing Director



Microsoft Research India
Inaugurated by

Shri Kapil Sibal
Honourable Minister of State
for Science and Technology
and Ocean Development

Government of India

Bangalore on 12th January, 2005

Why Did We Set up a Lab in India?

- Strong engineering education, talent pool
- Growing economy and aspirations for R&D and innovation
- Interesting research topics stimulated by the socio-economic-cultural environment

People & Place

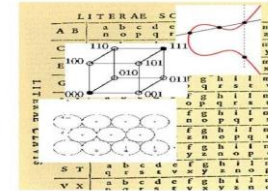
- Full-time staff total: 53
- Technical staff: 48
 - 34 with PhD (69%)
- Total interns (6 years): >500
 - Indian universities: 60%
 - Foreign universities: 40%
 - PhD candidates: 43%
 - Bachelors/Masters: 57%
- Universities include...
 - IISc, IITs, BITS, DAIICT, IIITs, MIT, CMU, UC Berkeley, UW, Yale, Cambridge, Chicago, Bristol, National University of Singapore, Institute of Mathematical Sciences, Oxford, Georgia Institute of Technology, Princeton, New York University, University of Maryland, McGill University, U. Cape Town, UC Santa Barbara, U Michigan



Groups

Research Groups

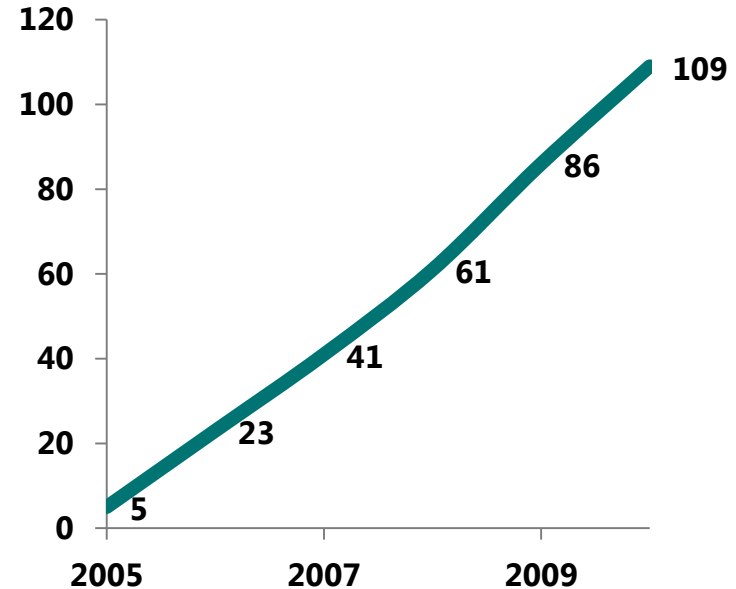
- Algorithms
- Cryptography, Security, and Applied Mathematics
- Mobility, Networks, and Systems
- Multilingual Systems
- Rigorous Software Engineering
- Technology for Emerging Markets
- Vision, Graphics, and Visualization
- +
 - Advanced Development
 - MSR Connections



Award Winning Research

- Best-paper awards in
 - ETAPS 2010 (RSE)
 - USENIX ATC 2010 (MNS)
 - CHI 2010 (TEM)
 - SECRIPT 2010 (CSAM)
- ACM SIGPLAN Outstanding Dissertation Awards 2010
 - Akash Lal
 - Bill Thies
- MIT TR35 2010
 - Indrani Medhi
 - Rikin Gandhi (alumnus)
- ACM Distinguished Members
 - Venkat Padmanabhan
 - Ram Ramjee
 - G Ramalingam
- New Jersey R&D Council Thomas Alva Edison Patent Award – Ram Ramjee
- EE and IDC Regional Innovation Award 2010 (DebugAdvisor)

No. of Papers Published per Year



Partner with Academia, Industry and Government

Strengthen CS Research Ecosystem

- Capacity Building
- Communitywide Research Initiatives

Address Societal Challenges

- Community Engagement
- Social Responsibility

Empower with Technologies

- Help advance research process
- Innovation

MRC Programs

- Events
- Summer and Winter Schools
- Theory Day
- Research Workshops
- MSR Faculty Summit
- TechVista
- Community Engagement
- Conference sponsorships
- RSE Tools Program (new)
- WikiBhasha & Worldwide Digital Heritage Community Programs (new)
- HPC and Science (new)
- Student programs
- PhD Fellowships
- Travel Grants
- Rx Lab
- Joint PhD Advising
- Raising Star Awards (new)
- Faculty programs
- Young Faculty Awards
- Research Collaborations & Grants
- Visiting Researcher



Microsoft® Research

FacultySummit 2011

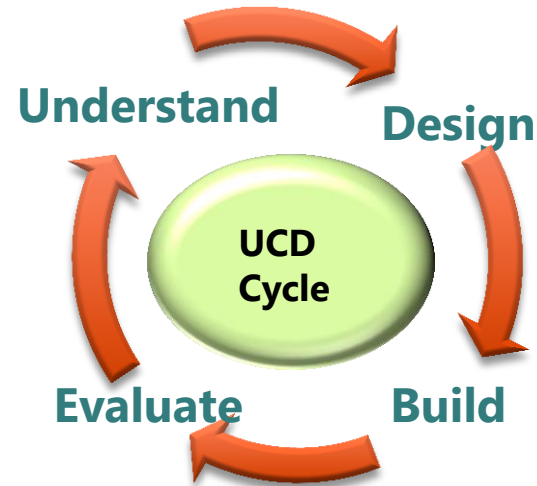
Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

ADDRESSING SOCIETAL CHALLENGES

User interfaces & technology for development (“HCI4D”)

- User-centered design
 - “Impedance match” of technology & people
 - Work in the field, studying people, their context, their needs and aspirations
 - Design & build accordingly
 - Does it do what you hoped?
 - Revise, repeat

“NUI” is often the only thing that makes sense



Paper, Pen & Digital Slate



“Natural” leverages
the habitual and the comfortable

Can handwritten pen and paper-based work processes directly generate and manipulate digital data?

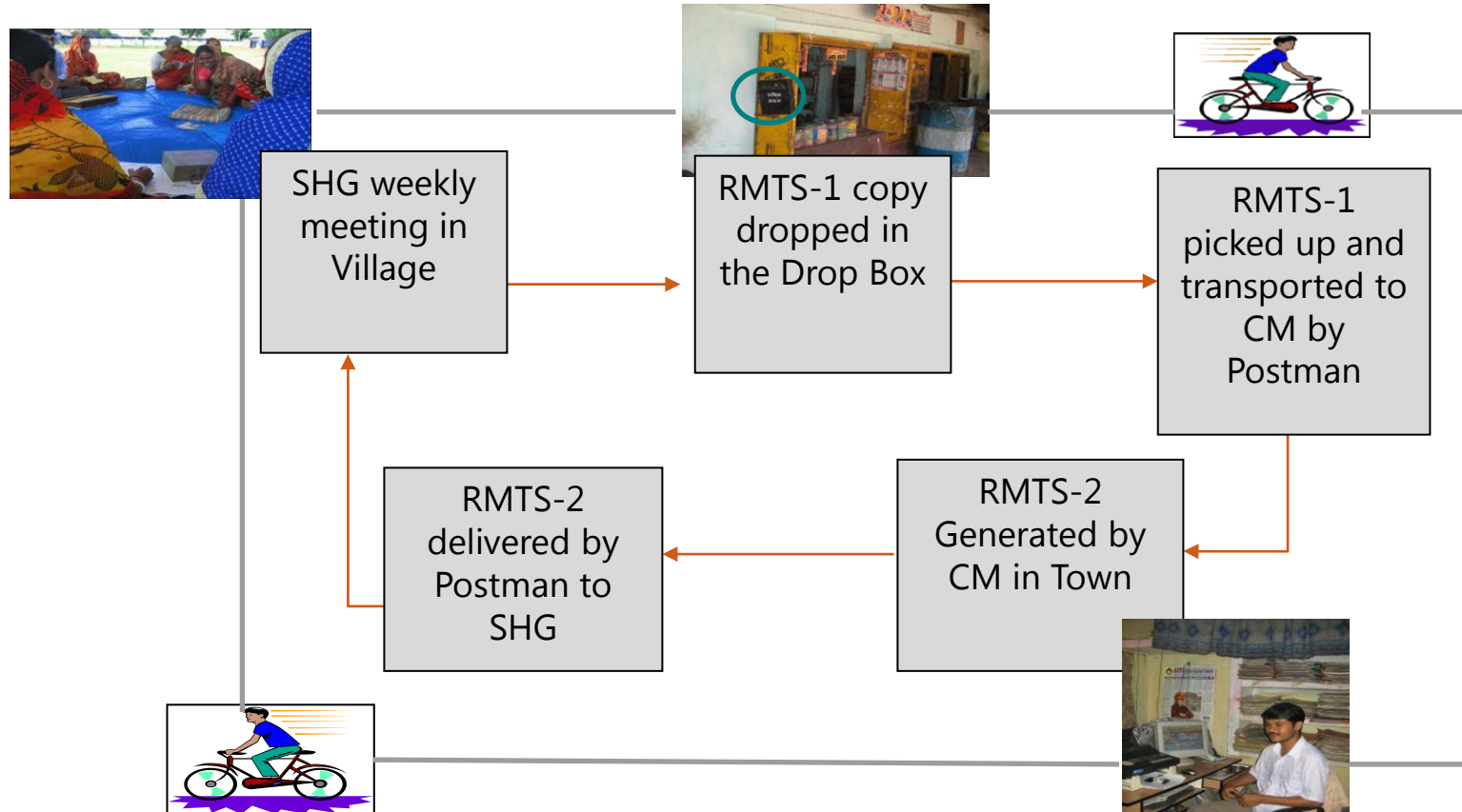


Domain

- 86 million women participants in 6 million microfinance Self-Help Groups across India
- Linked with banks
- Decentralised, autonomous, self-run
- Active in leading social, political and economic initiatives, but limited financial leverage/growth [Sinha et al, EDA & APMAS, 2006]



PRADAN'S "Computer Munshi"



Key issues in baseline system

সাপ্তাহিক বৈঠকে লেন-দেন বিবরণী 43

তারিখ : 15.11.19 গ্রুপের নাম : পাইলট/খি 3/1 অর্ডিনে 3 গ্রাম : বেড়া, 1/1 নিচিং সংখ্যা : 381

ক্রমিক সংখ্যা	সমন্বায়ক নাম	ক্রেত সংখ্যা	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট	ক্রেত মোট
১	আব্দুল হক	✓	70															
২	আব্দুল হক	✓	20															
৩	আব্দুল হক	✓	10															
৪	আব্দুল হক	✓	5															
৫	আব্দুল হক	✓	2.5															
৬	আব্দুল হক	✓	5															
৭	আব্দুল হক	✓	10															
৮	আব্দুল হক	✓	28															
৯	আব্দুল হক	✓	15															
১০	আব্দুল হক	✓	10															
১১	আব্দুল হক	✓	20															
১২	আব্দুল হক	✓	20															
১৩	আব্দুল হক	✓	30															
১৪	আব্দুল হক	✓	3000	3000														
১৫	আব্দুল হক	✓	100															
১৬	আব্দুল হক	✓	10															
১৭	আব্দুল হক	✓	23															
১৮																		
১৯																		
২০			411.5			300	250											

গ্রুপের আয় : 411.5 গ্রুপের ব্যয় : 250 মোট আয় : 161.5

পূর্বে মোট ছিল 608.10
মোট আয় : 161.5 মোট ব্যয় : 366.11

শেষের সই : *PRADAN* সভাপতির সই : *Sunandan Chakraborty* বাকী জমা (ক - খ) : 4258.5

ক্যাশ ট্যাগ	১	২	৩	৪	৫	৬	৭	৮	৯	১০	১১	১২	১৩	১৪	১৫	১৬	১৭	১৮	১৯	২০	
১০০ X																					
১০০ X																					
১০ X																					
২০ X																					
১০ X																					
৫ X																					
২ X																					
১ X																					
০.৫০ X																					
মোট																					

Data Accuracy

- Recording Errors
- Calculation Errors
- Legibility Errors

Data Completeness

Process Efficiency

- Error Resolution
- Paper Form Transport Delays
- Extended Meeting Time

How Does It Work?

Simultaneous
digitization of entries

Process entries against
database; prompts and checks

Handwriting
numbers on paper

Audio report
for validation



Field Trials

Within subjects comparison

Trial 1: West Bengal, 10 SHGs, median education: 7 yrs (w/ munshi)

Trial 2: Orissa, 4 SHGs, median education: 3 yrs (no munshi)



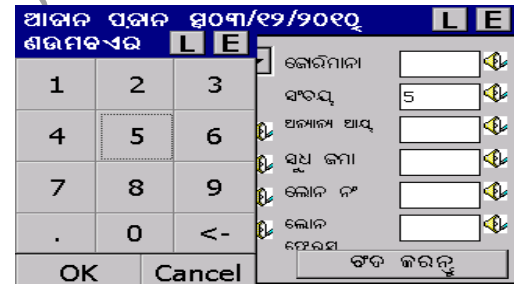
PAPER-ONLY (PO)

Baseline system
Trial 1 and 2



DIGITAL SLATE (DS)

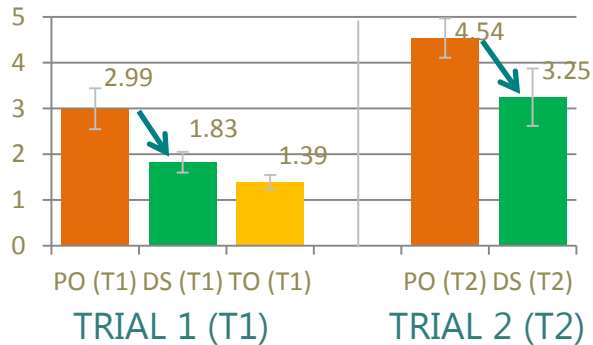
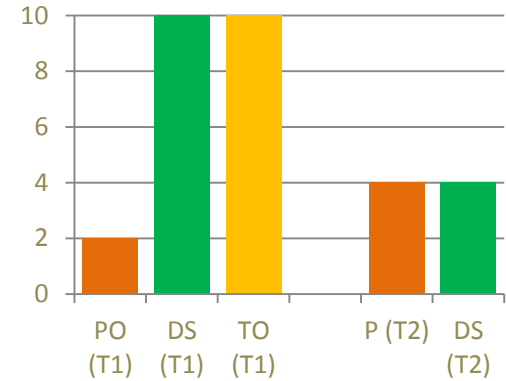
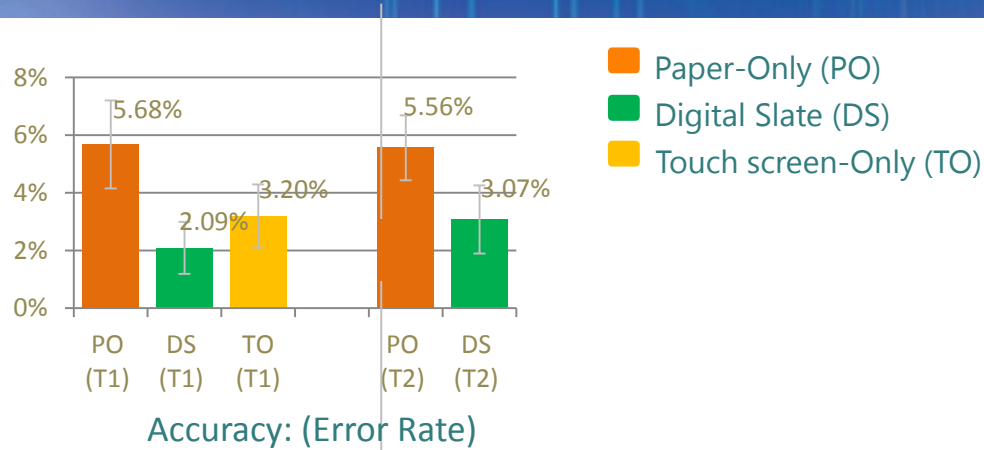
Designed solution
Trial 1 and 2



TOUCH-SCREEN ONLY (TO)

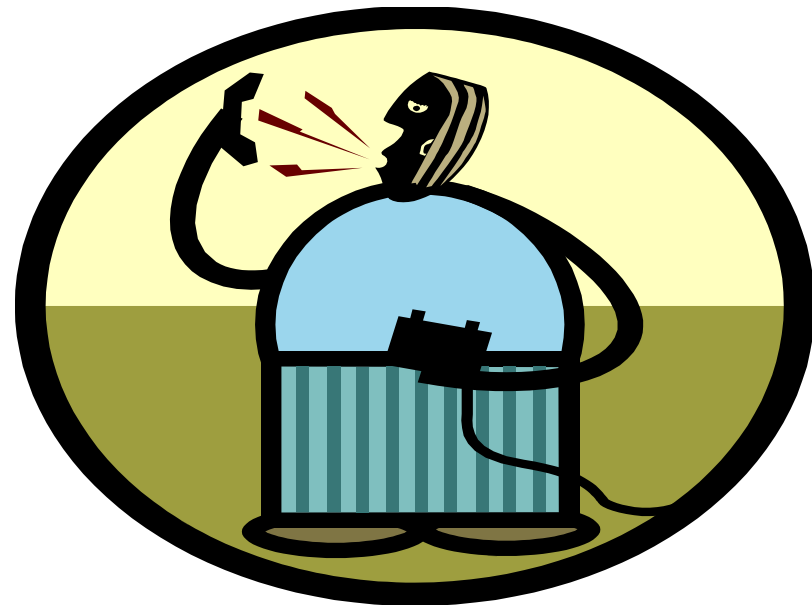
Alternative solution
Trial 1

Field Trial Results



Efficiency: meeting time

From Paper to Voice: Natural Modes of Communication

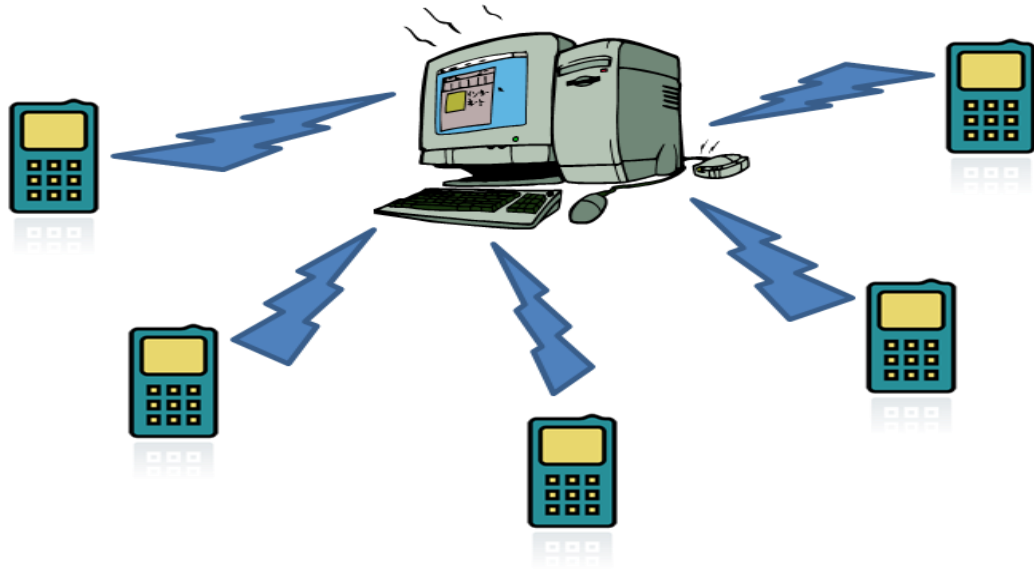


The Power of the Phone

- Internet penetration in India was 6.9% in 2010
 - For the state of Chhattisgarh, this number is 0.7%
 - More problems for Web in Hindi (not to mention other languages)
- Issues like power, education and finances limit reach of Web
- In contrast, mobile penetration in India stands at 60% and growing rapidly
- Cheap instruments, cheap call rates, oral medium
 - Cost matters, and voice is the natural communication medium
- Audio transcends the language-literacy barrier



IS IT POSSIBLE TO BRING THE POWER OF SOCIAL MEDIA TO 800 MILLION RURAL INDIANS VIA MOBILE PHONE?



CGNet Swara

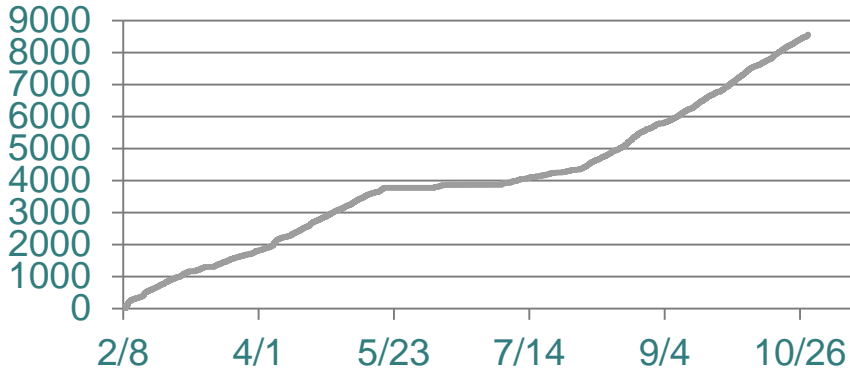
Voice Portal for Citizen Journalism

- Anyone can report news, issues, etc. in local language
- Submissions are reviewed by moderators over the Web
- Appropriate submissions are published:
 - For playback on audio channel
 - For browsing on web
 - Some submissions seed stories for posting on CGNet site + list

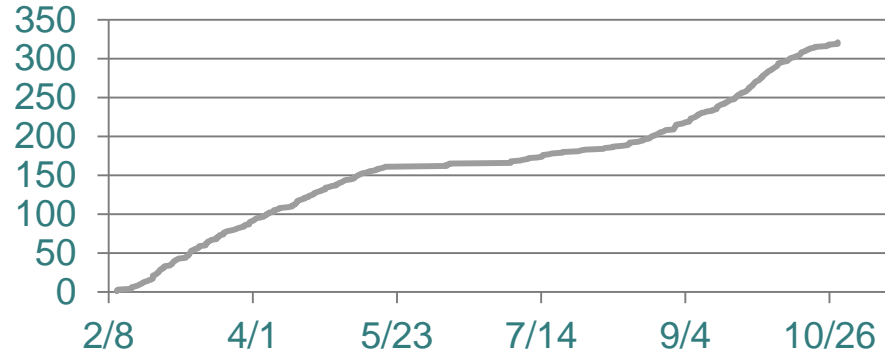


Deployment – The first 8 months

Calls



Published Reports



Who is reporting?

- Mostly activists/social workers
- Handful of trained journalists
- Majority did not attend training session

Languages used:

- Hindi
- Chhattisgarhi
- Gondi
- Kurukh

Application – Governance and Grievance Redress, News and Songs

- About 40% of all posts concerns with governance
 - Government employment scheme (NREGA) workers not getting paid
 - Forest land not distributed under FRA
 - Situations of rural schools and health centers
 - Reports on corruption in public schemes
 - Teachers not being paid
 - The teachers were paid their back wages one week after publication of the interview on CGNet Swara
- Real time breaking news directly from the source in local languages
 - Reports on pollution and public health emergencies
 - Deaths due to cholera, diarrhea and malaria
 - Arsenic in water
- Tribal stories and songs

Application – News and Others

- Several real time breaking news directly from the source
- News in Local Languages
 - First and Only source of news in Gondi and Kurukh
- Reports on pollution and public health emergencies
 - Deaths due to cholera, diarrhea and malaria
 - Arsenic in water
- Alternative narrative of the same event
 - Report on a public hearing
- Live reports on tribal festivals
- Oral traditions recorded directly from the source
 - Tribal stories and songs





Microsoft® Research

FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

WIKIBHASHA

WikiBhasha



The screenshot shows the WikiBhasha Beta website. At the top left is the "WikiBhasha" logo with "Beta" in a smaller font above it. To the right is the "Microsoft Research" logo. Below the logo is a navigation bar with links: Home, Install, Take a Tour, FAQ, Discuss, Developers, About, and a Share button with Facebook and Twitter icons. The main content area has a heading "WikiBhasha Beta" followed by a paragraph describing the tool as a multilingual content creation tool for Wikipedia, developed by Microsoft Research. It mentions that it enables users to explore and source content from English Wikipedia articles, translate it into target languages, and use it with user additions and corrections. A second paragraph highlights its features: easy-to-use interface, quick translation of articles from English to over 30 other languages, and support for Microsoft's Machine Translation system and Collaborative Translation Framework. Below this is a call to action: "Take WikiBhasha beta for a test drive by clicking on the install link" with a link to "click here to read the W...". A large "Install" button is prominent, with the text "Click here to install WikiBhasha beta". At the bottom left, there are links for "Microsoft Research | Website Terms of Use | Privacy S...".



Multilingual content creation tool for Wikipedia

Also enables community creation of parallel corpora for language research

In collaboration with Wikipedia

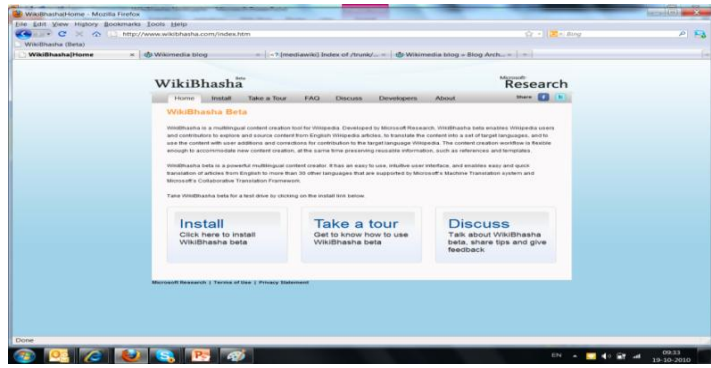
Open source tool

Launched in October 2010

Purpose (community & research impact)

- ◆ A collaborative tool for Multilingual Wikipedia content creation
 - ◆ Machine Translation and Collaboration Mechanism as resources
- ◆ Creation of Linguistic Parallel data for research

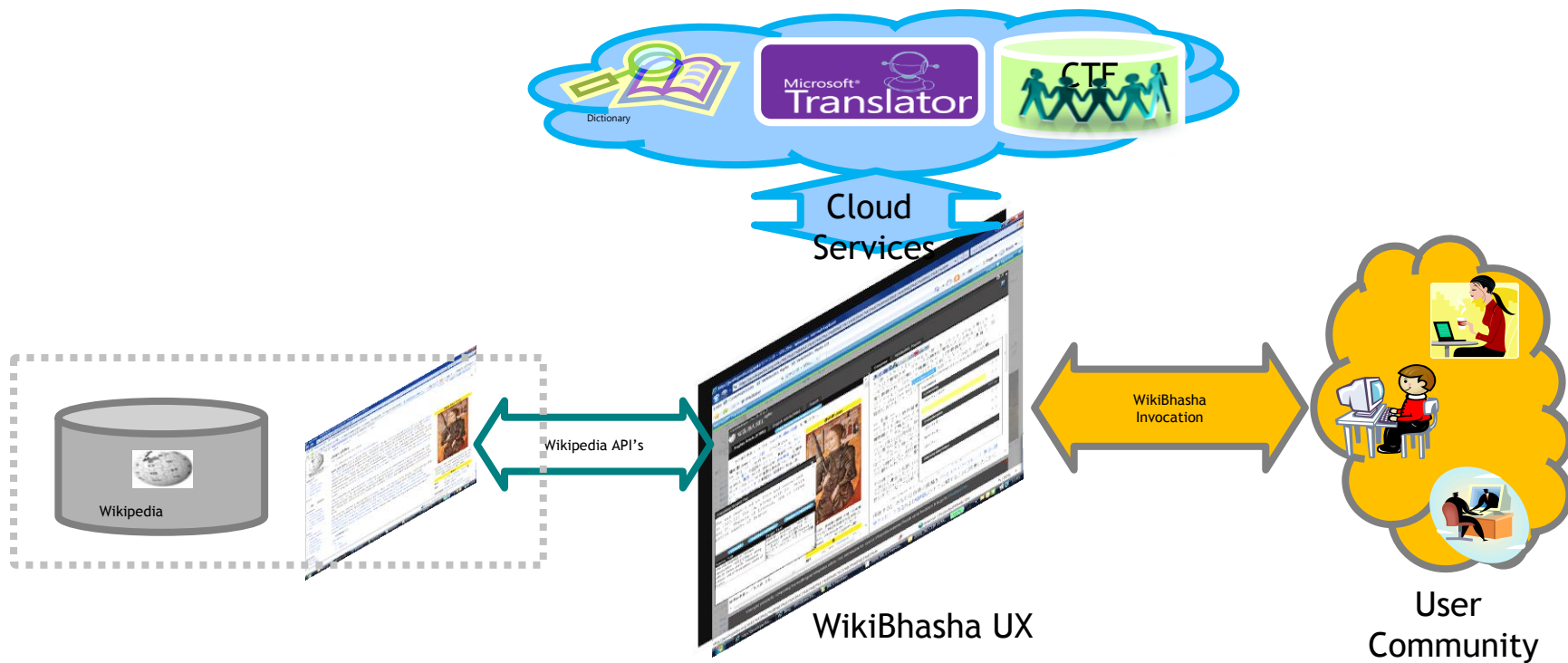
Release as a open-source tool



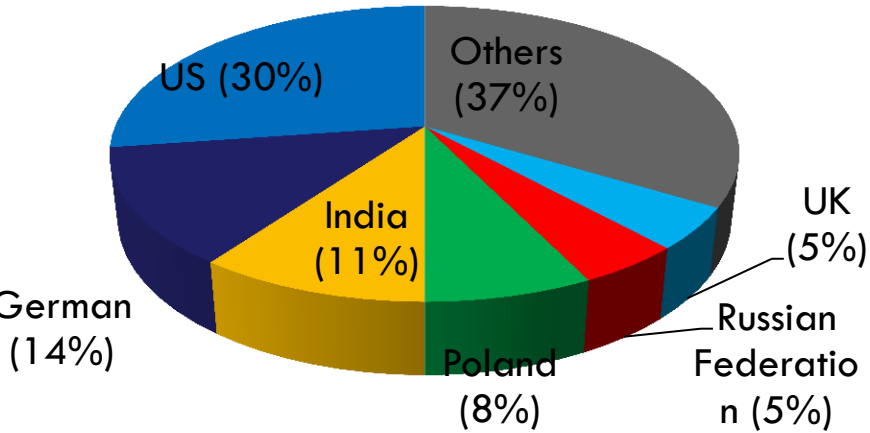
WikiBhasha released as a open-source tool:

- As MediaWiki Extension in <http://svn.wikimedia.org/viewvc/mediawiki/trunk/extensions/WikiBhasha/>
- As Wikipedia User-script from WikiBhasha.MSR user
- As Bookmarklet in www.WikiBhasha.org

Approach



Plans



- Working with Wikimedia Foundation & Wikipedia Communities
- Workshops planned for
 - Hindi
 - Spanish
 - Brazilian Portuguese
 - Arabic





Microsoft® Research

FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

RICH INTERACTIVE NARRATIVES

Digital Heritage – Two Modes of User Experiences

Explanatory Narratives

- Text documents, video narratives, audio
- Linear, Authored, Informative and Explanatory
- Author in control of user attention
- User gains knowledge in a guided fashion
- High production effort
- Informative content

Interactive Exploration

- Image Galleries, hyperlinked documents, 3D walkthroughs
- Non-linear browsing
- User in control of her attention
- User gains knowledge in an ad-hoc fashion
- Low production effort
- Often no annotation or explanation of content is provided

*Narratives **explain** but limit the audience*

*Interactive **exploration** is liberating, but not informative*

Rich Interactive Narratives (RIN)

Combine the power of narration and exploration

Enable seamless integration of a variety of linear and interactive media

Provide an easy tool for authoring content

Deliver content via the browser (no need for client app)

Create an extensible framework for adding new media and new types of experiences

digital narratives

EXPLORE PEOPLE FAQ BLOG

FEATURED



LADAKH - THE DOMKHAR VALLEY

A snapshot of a traditional valley in Ladakh in the midst of modernizing influences. Observed over successive visits to the region in the winter and summer of 2010 by a group of four travellers from Microsoft Research. Special thanks to Cynthia Hunt, long time resident of Ladakh for hosting us and making this narrative possible.

This is a preview of the Rich Interactive Narratives technology created by Microsoft® Research. The narratives require Microsoft® Silverlight™ version 4, which is available [here](#). The narratives are best experienced in full-screen mode with an internet connection speed of 2 Mbps or higher.

WHAT IS A DIGITAL NARRATIVE?

A digital narrative is a narrative that takes advantage of multimedia and web technologies, providing the author with a new way to build and present a media-rich story, and simultaneously offering the viewer a new way to interact with the content. We, at Microsoft® Research

WHAT DOES "INTERACTION WITH A NARRATIVE" MEAN?

Typically, viewing of a traditional narrative such as a movie has been a passive experience, where the viewer is limited to the path the content author has defined. The

THE RICH INTERACTIVE NARRATIVE TECHNOLOGY

One of the most critical tasks in creating a new media format and narration method is to define an effective data model. We have defined, in the course of this





Photo:
Paige Bills

Thank you!