

Offline Wikipedia on Low-Cost Devices

Bill Thies

Joint work with Gaurav Paruthi and Arunesh Mathur

Microsoft Research India

Established January, 2005

Seven **research** areas

- Algorithms
- Cryptography, Security & Applied Math
- Graphics and Visualization
- Mobility, Networks, and Systems
- Multilingual Systems
- Rigorous Software Engineering
- *Technology for Emerging Markets*

Contributions to Microsoft:

- WikiBhasha, MultiPint, Virtual India

Currently ~55 full-time staff

Collaborations with government, academia, industry, and NGOs in India

<http://research.microsoft.com/india>



Microsoft Research India
Lavelle Road, Bangalore

Technologies for Emerging Markets



Photo: Divya Ramachandran

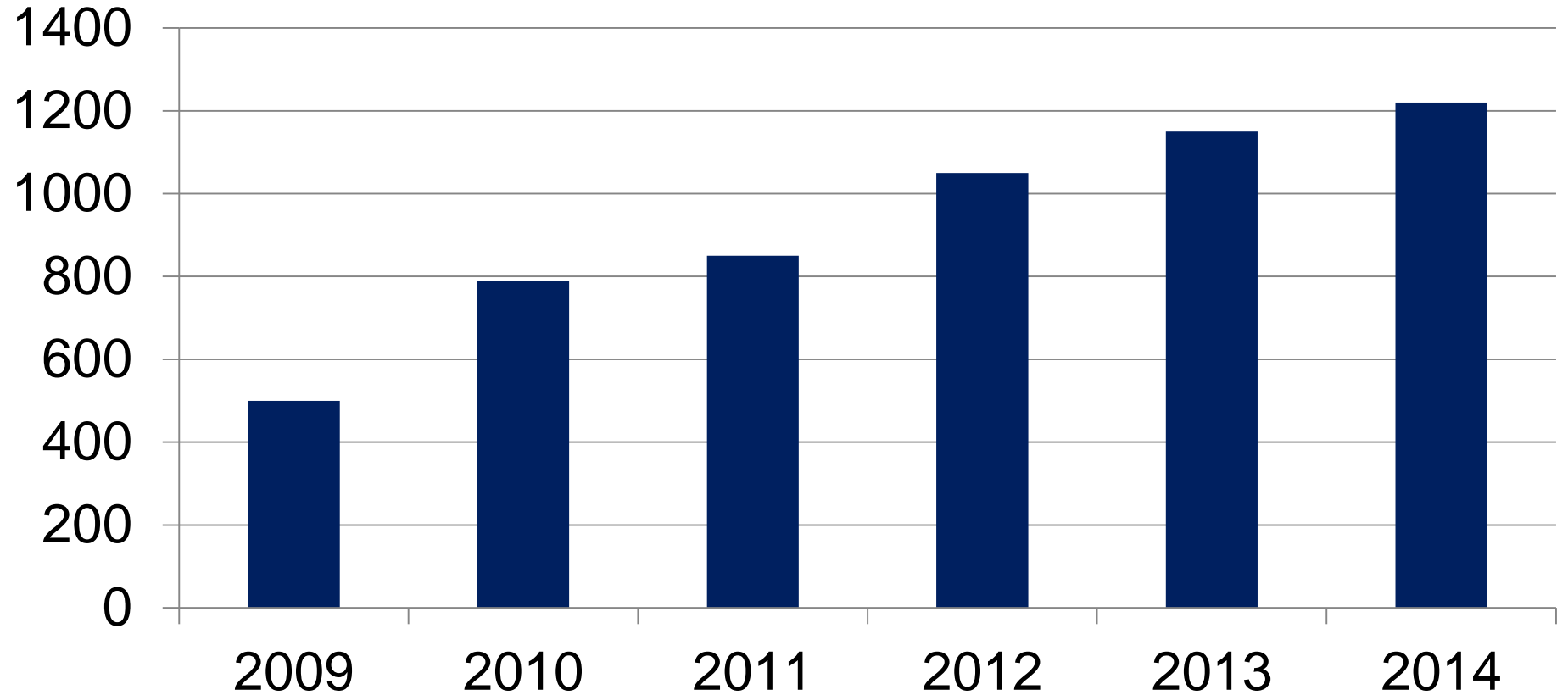
Understand potential technology users in developing communities

Design and evaluate technology and systems that contribute to socio-economic development of poor communities worldwide

Collaborate with development-focused organizations for sustained, scaled impact

Part 1:
Offline Wikipedia on Low-End Phones

India's Wireless Subscriber Base (millions of subscribers)



India's Wireless Subscriber Base

- Most subscribers lack smart phones

Smart Phone: < 5%



Feature Phone: ~50%
(e.g., music player)



Basic Phone: ~45%



Mobile Internet: < 3% of subscribers

Prior Solutions Target Smart Phones

- **Wikipock** (for Windows mobile, Blackberry, Android, Symbian S60)
- **WikiOnBoard** (for Symbian S60, Maemo, Linux and Windows mobile)
- **WikiReader** (OpenMoko's custom \$99 device)
- **Okawix** (For Android, Windows mobile)
- **Evopedia** (Cancelled for Symbian, Maemo)
- **Mopedi** (For touchscreen Nokia phones only)

 **Can we design a solution for feature phones?**

Offline Wikipedia for Feature Phones

- **Map schools-wikipedia.org to a feature phone**
 - Selected subset of ~5,500 articles, for education
- **Technical requirements:**
 - Java J2ME programmability (very common)
 - 2 GB SD card

NOKIA

11:54
Wikipedia 24

MUMBAI

Options Ok Clear

NOKIA

Mumbai



Menu Back

NOKIA

Mumbai

of the Indian state of Maharashtra. It is the most populous city in India, and the sixth most populous city in the world, with a population of approximately 20.5 million. Along with the neighbouring urban areas, including the cities of Navi Mumbai and Thane, it is one of the most populous urban

Menu Back

Formatting Pages for Feature Phones

- **Start with** m.wikipedia.org content.
- **Remove tags** incompatible with mobile, then **paginate**:

The screenshot shows a mobile version of the Wikipedia article for 'Mantra-Rock Dance'. A large red oval highlights the top section of the page, including the title, a poster image, and the introductory paragraph. Below this, a list of expandable sections is shown, with red ovals around the 'Background', 'Preparation and promotion', 'Event', 'Reaction and effect', and 'External links' buttons. Arrows point from these sections to labels on the right: 'Introduction' (pointing to the top section), 'Background' (pointing to the 'Background' section), 'Preparation & promotion' (pointing to the 'Preparation and promotion' section), 'Event' (pointing to the 'Event' section), and 'Reaction & effect' (pointing to the 'Reaction and effect' section). A red oval around the 'External links' button points to the label 'Ignored'. At the bottom, a link says 'View this page on regular Wikipedia'.

Introduction

Background

Preparation & promotion

Event

Reaction & effect

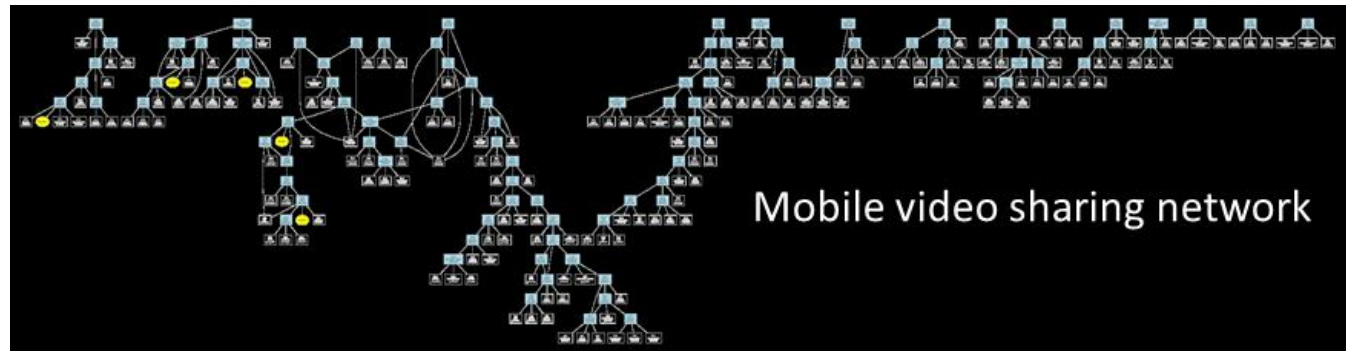
Ignored

Technical Details

- **Java J2ME phones do not have built-in browser**
 - Identified and adapted Fire-J2ME browser
- **Compression and cache management**
 - Every page is compressed separately (DEFLATE)
 - Compressed pages archived into single file, to overcome limitations on # files per directory
 - Upon visiting a page, it is uncompressed into a 1 MB cache on SD card (using LRU replacement policy)
- **Search using pre-built alphabetical index**
 - Populate index with all permutations of words in title
 - Modify HTML links to point directly to target index

Deployment Strategy

- **Sell in mobile shops!**
 - Alongside movies, music, etc., on SD cards
 - Potential to select pages based on users' interests
- **Transfer peer-to-peer using Bluetooth**



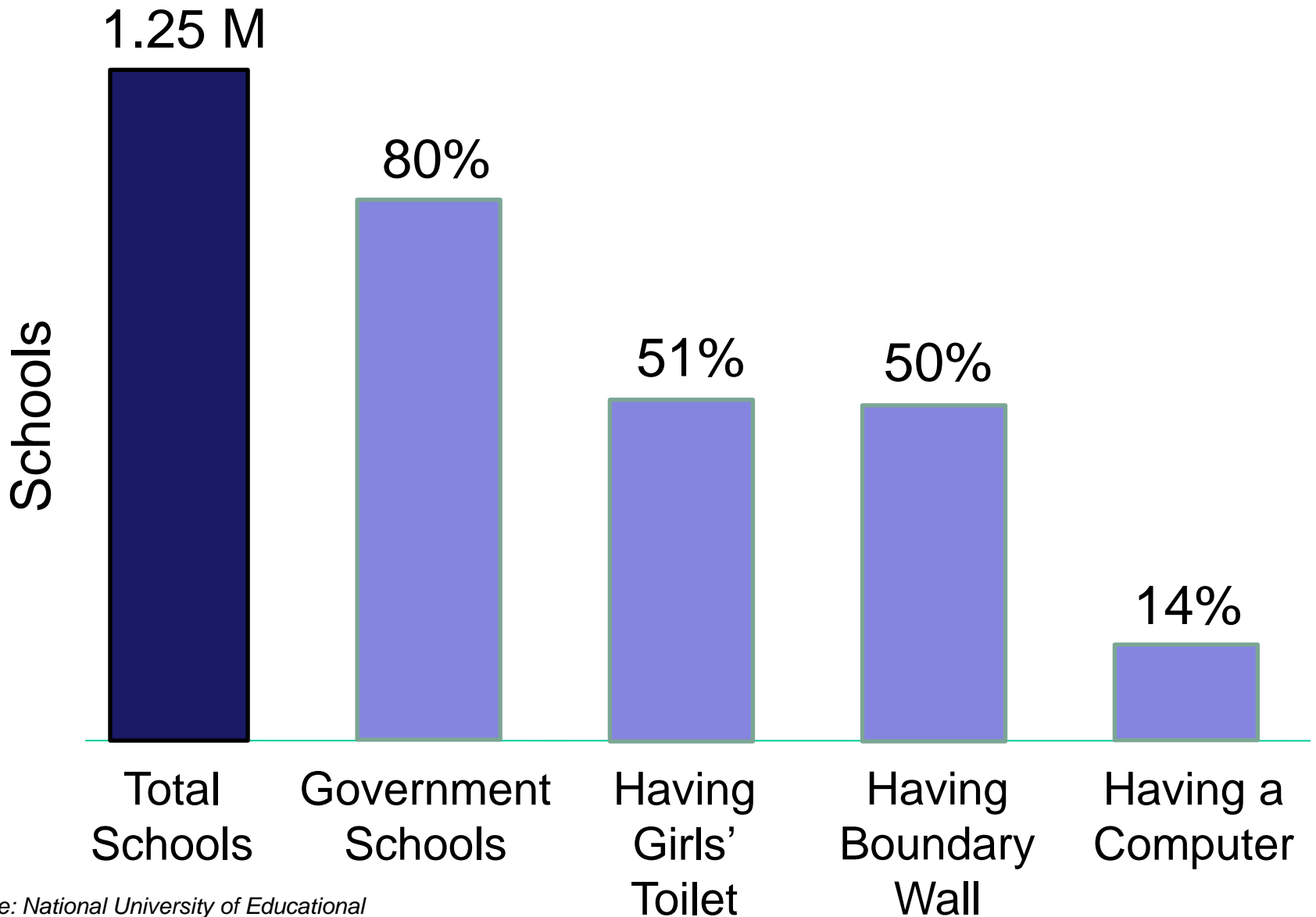
Credit: Thomas Smyth

How to Update the Content?

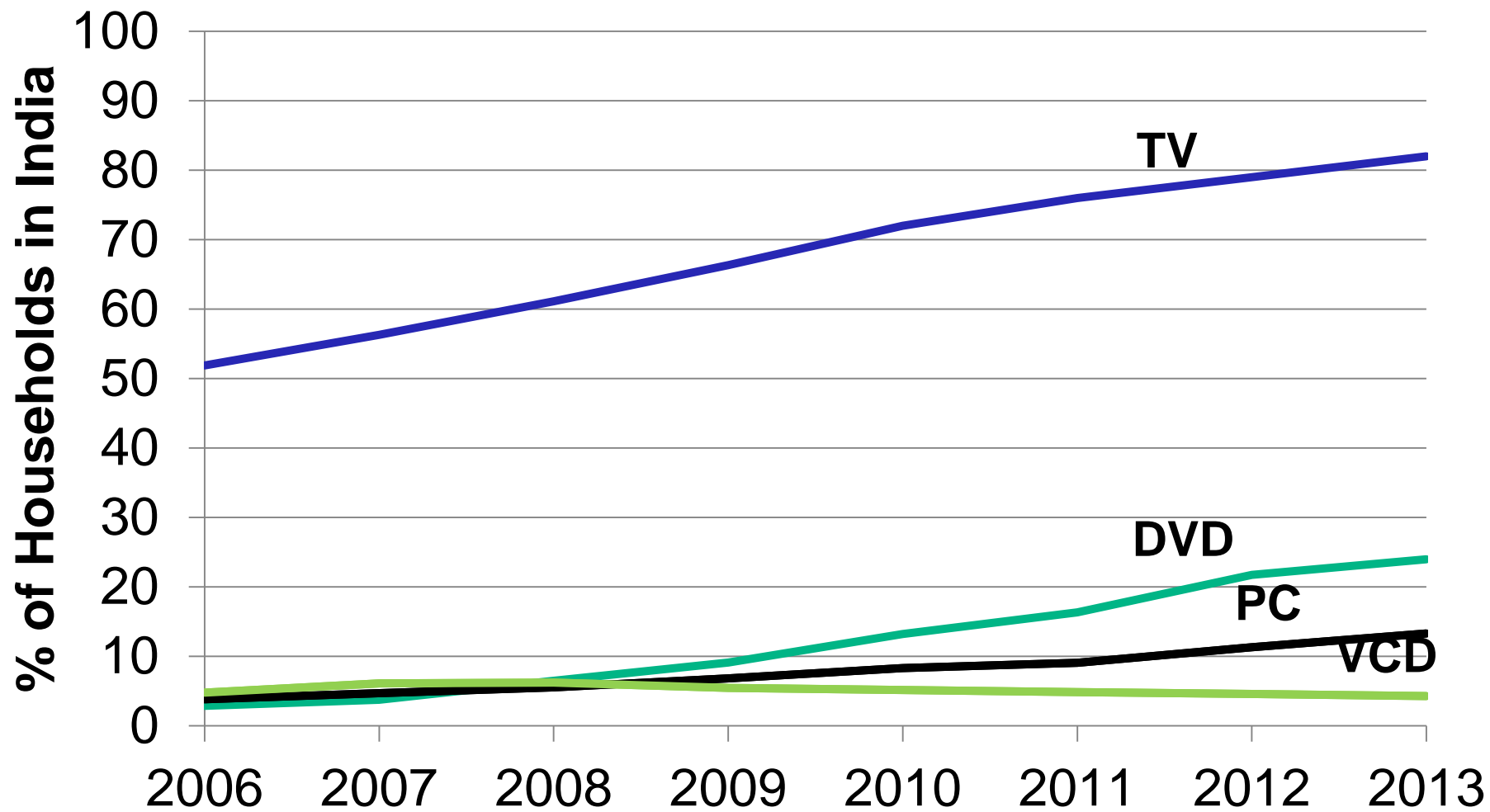
- **Periodically republish curated updates**
- **Allow contributions from users via SMS**

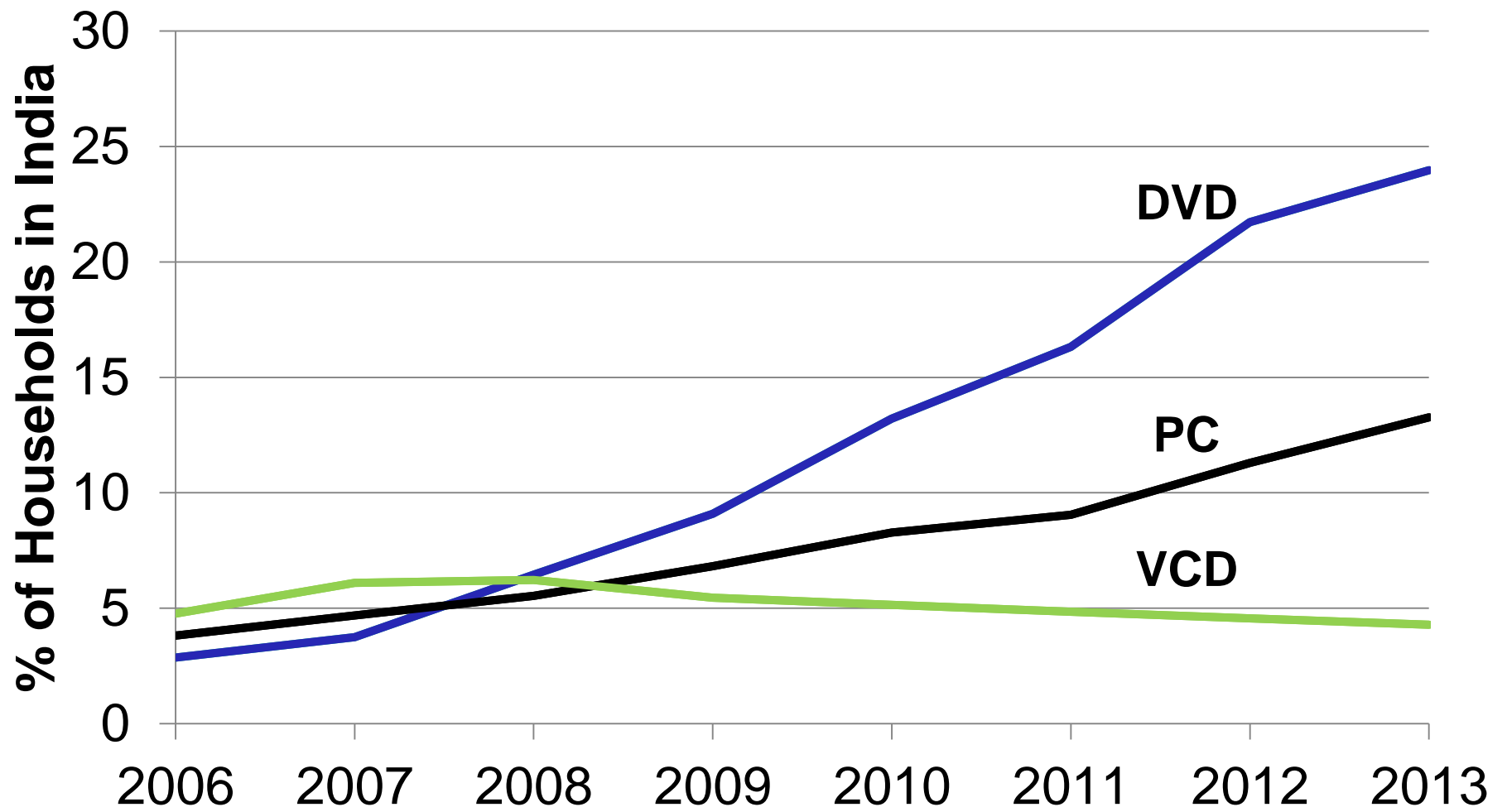
Part 2:
Offline Wikipedia on TV-DVD

Classroom Environment in India











Some photos courtesy Nithya Sambasivan

Schools-Wikipedia on TV-DVD

- Mapped schools-wikipedia.org to a DVD
 - 5,500 articles
 - 259,000 screens on TV
- Interactive DVDs can imitate most functions of web browser:
 - Search
 - Subject/title index
 - Hyperlinks
 - Back button
- As far as we know, this is the **cheapest means to distribute Wikipedia**
 - Cost of DVD: \$0.50
 - Cost of TV/DVD Player: \$50 (10" B&W TV)



① Search

② Subject Index

③ Title Index





Search

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

DONE





VAN

A

B

C

D

G

I

U

DONE

BKSP

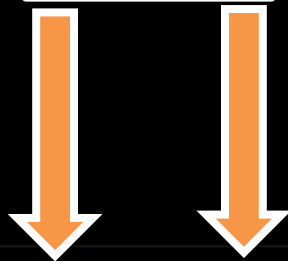




- van Beethoven, Ludwig
- Vanbrugh, John
- Van Buren, Martin
- 8** Vancouver
- 9 Van der Waals force
- 10 van Dyck, Anthony
- 11 van Gogh, Vincent
- 12 Vanguard, Project
- 13 Vanilla
- 14 Vanuatu
- 15 vapor, Water
- 16 Varanasi



Hyperlinks



Navigation Buttons

Hyperlink



Area

Units: **5** Mathematics
Area is a **6** quantity expressing the two-dimensional size of a defined part of a surface, typically a region bounded by a closed **7** curve. The term surface area refers to the total area of the exposed surface of a 3-dimensional solid, such as the sum of the areas of the exposed sides of a **8** polyhedron.

Units

Units for measuring surface area include:



User Study



Wikipedia on Computer

VS.

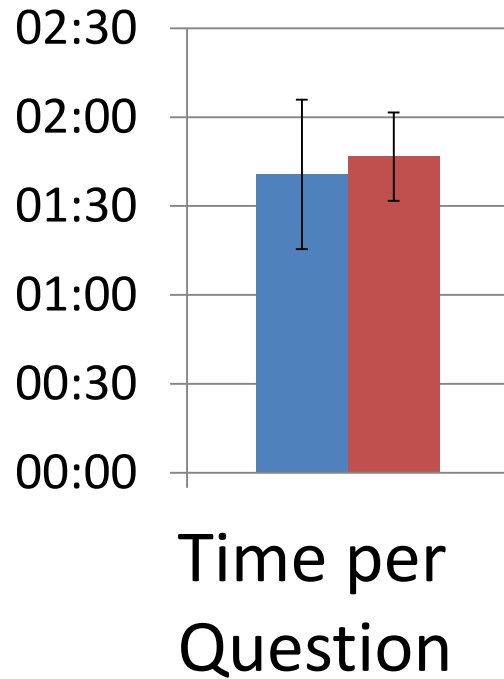
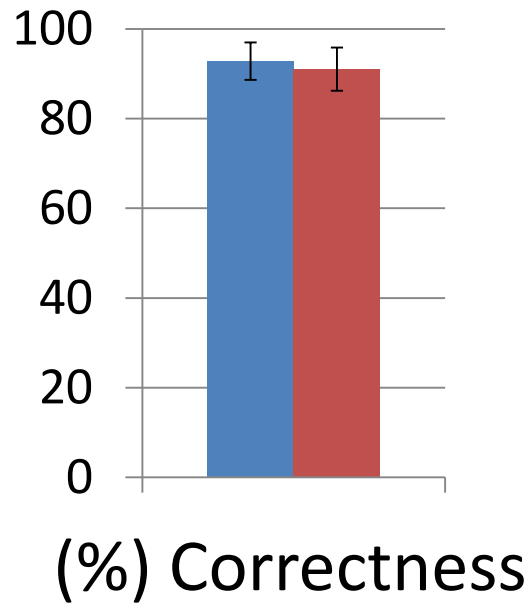


Wikipedia on TV-DVD

- **User group:**
 - 20 low-income college students & office staff in Bangalore, India
 - None own a computer, but 75% own TV-DVD
- **Task: Answer 10 questions (half computer, half TV-DVD)**
 - What is the capital of Ukraine? Who discovered Pluto?

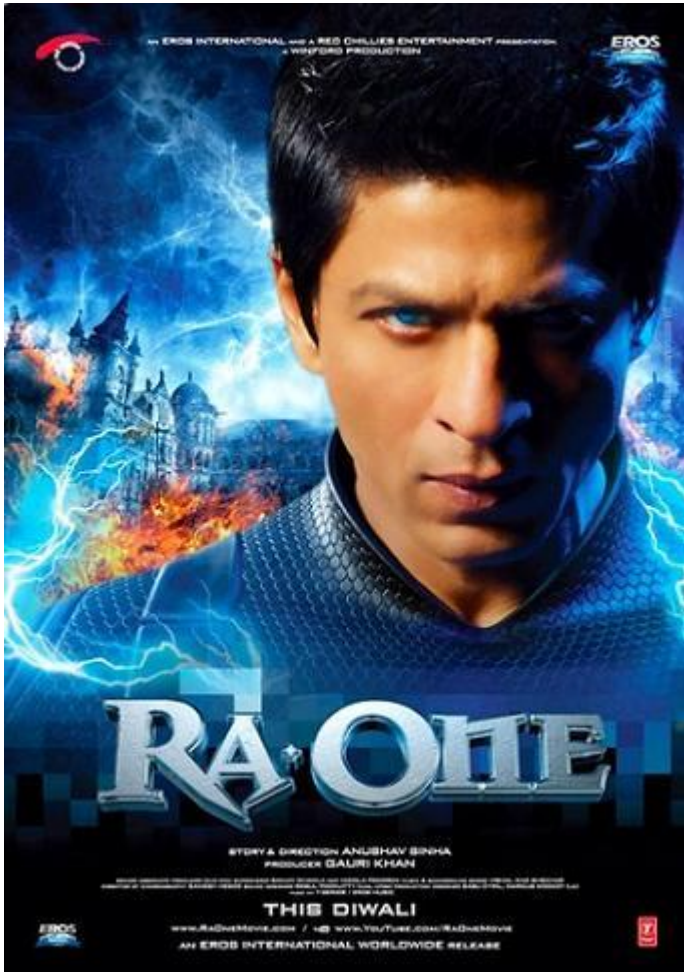
User Study

Computer TV-DVD



Deployment Strategy

- Include Wikipedia subset on Bollywood DVDs?



+

Ra.One

From Wikipedia, the free encyclopedia

This article is about the Hindi motion picture. For the super (character).

Ra.One (Hindi: रा.वन; full: *Random Access–Version 1.0*) is a 2011 Indian science fiction superhero film written and directed by Anubhav Sinha. The film features Shahrukh Khan in dual roles, and also stars Kareena Kapoor, Armaan Verma and Arjun Rampal in the lead. Shahana Goswami, Dalip Tahil and Chinese-American actor Tom Wu appear in supporting roles, along with Rajinikanth, Sanjay Dutt and Priyanka Chopra making guest appearances. The film follows Shekhar (Khan), a goofy game designer living and based in London along with his wife (Kapoor) and son (Verma), who designs a motion-sensor based game in which the villain is far more powerful than the hero. In the process, the villain escapes from the video game and begins to track down the only person ever to defeat him, Lucifer (Verma). *Ra.One* was jointly produced by Eros International and Khan's production company, Red Chillies Entertainment.^[3] Initially expected to hit theatres on June 3,

Conclusions:

Offline Wikipedia on Low-Cost Devices

- **In India, large demand for information from users lacking computer and Internet access**
- **Two complementary solutions for offline Wikipedia:**
 - On Java J2ME phones: small display, personal usage
 - On TV-DVD: large display, shared usage
- **Both in prototype stage, hoping for open public release in 2012**