BUILDING KNOWLEDGE BASES AND PLATFORMS VIA MASS COLLABORATION ON THE INTERNET

Institute on Internet and Society

June 9, 2013

T. Vishnu Vardhan

Prg. Director Access to Knowledge, Centre for Internet & Society

✉ vishnu@cis-india.org   vishnuvardhan_t
What was a “Knowledge Platform” like ................ in

2000
1950
1850
1650
1450

2050 ?? ????

vishnu@cis-india.org
• What does Print Technology tell us?
• Techno-sociaity of Knowledge production
• Tectonic Shifts in the Knowledge domain

Production-Dissemination-Distribution-Constitution

vishnu@cis-india.org vishnuvardhan_t
• What are the changes that we see?

Caxton Showing the First Specimen of His Printing to King Edward IV at the Almonry, Westminster circa 1851.
All images in this PPT otherwise mentioned are from commons.wikimedia.org See the last slide for credits

✉ vishnu@cis-india.org  
🐦 vishnuvardhan_t
History of books
From Wikipedia, the free encyclopedia

This article may require cleanup to meet Wikipedia's quality standards. No cleanup reason has been specified. Please help improve this article if you can. January 2009

The history of books follows a suite of technological innovations for books. These improved the quality of text conservation, the access to information, portability, and the cost of production. This history has been linked to political and economical contingencies, the history of ideas, and the history of religion.

Contents[show]
1 Origins and antiquity
2 Clay tablets
3 Papyrus
4 First books
5 Pre-Columbian codices of the Americas
6 Wax tablets
7 parchment
8 Paper
9 Middle Ages
9.1 Books in monasteries
9.2 Copying and consuming books
Some Massive Knowledge Platforms

- Baidu Baike
- English Wikipedia
- Hudong
- Catawiki
- WikiEducator
- Open Street Map
- Pad.ma
- Sahapedia
- Internet Archive
- JSTOR
- DSAL
- DLI
• The spread of the internet has made possible the building of knowledge bases by seamless and mass collaboration. But HOW?
WIKI

• 2 minute History
  – 1994 - Ward Cunningham

• Simple Video on Wiki 🎥

• Browser based editing & version history

✉ vishnu@cis-india.org
• 286 Language Wikipedias
• 334 Language Wikipedias in incubation
• Total Articles – 2,60,00,000
• Articles in English Wikipedia – 42,00,000
• 16,43,063 Editors/Contributors
• 73,049 Active Editors/Contributors
• 36,50,00,000 Page Views per month
• No. of paid employees..............................173

✉ vishnu@cis-india.org

vishnuvardhan_t
Top 5 Indian Wikipedias have 2,56,160 Articles
Total Indian Contributors – 5514
Very Active Editors – 467
2,75,79,239 page views per month

## Indian Language Wikipedia - Factsheet
Indian Language Wikipedia Statistics - as on March 31, 2013

<table>
<thead>
<tr>
<th>Language</th>
<th>Hindi</th>
<th>Bengali</th>
<th>Punjabi</th>
<th>Marathi</th>
<th>Telugu</th>
<th>Tamil</th>
<th>Urdu</th>
<th>Kannada</th>
<th>Gujarati</th>
<th>Malayalam</th>
<th>Oriya</th>
<th>Assamese</th>
<th>Kashmiri</th>
<th>Sanskrit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Page Views per Month</td>
<td>5,926,278</td>
<td>3,028,411</td>
<td>344,503</td>
<td>2,913,713</td>
<td>1,456,140</td>
<td>3,434,075</td>
<td>1,291,857</td>
<td>891,997</td>
<td>668,153</td>
<td>2,294,413</td>
<td>292,777</td>
<td>349,024</td>
<td>54,421</td>
<td>383,750</td>
</tr>
<tr>
<td>Article Count</td>
<td>97,151</td>
<td>25,725</td>
<td>5,690</td>
<td>39,313</td>
<td>52,522</td>
<td>52,431</td>
<td>23,670</td>
<td>14,255</td>
<td>22,611</td>
<td>30,250</td>
<td>3,690</td>
<td>1,990</td>
<td>113</td>
<td>10,272</td>
</tr>
<tr>
<td>New Articles per Day</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>8</td>
<td>21</td>
<td>13</td>
<td>11</td>
<td>2</td>
<td>23</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Edits per Month</td>
<td>33,874</td>
<td>29,531</td>
<td>5,720</td>
<td>8,758</td>
<td>16,238</td>
<td>47,551</td>
<td>25,450</td>
<td>10,113</td>
<td>12,832</td>
<td>29,077</td>
<td>4,142</td>
<td>2,485</td>
<td>132</td>
<td>6,306</td>
</tr>
<tr>
<td>Active Editors</td>
<td>44</td>
<td>61</td>
<td>10</td>
<td>29</td>
<td>36</td>
<td>86</td>
<td>26</td>
<td>25</td>
<td>20</td>
<td>124</td>
<td>13</td>
<td>14</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>Very Active Editors</td>
<td>9</td>
<td>8</td>
<td>3</td>
<td>8</td>
<td>12</td>
<td>23</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>26</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>New Editors</td>
<td>14</td>
<td>17</td>
<td>3</td>
<td>8</td>
<td>6</td>
<td>20</td>
<td>4</td>
<td>12</td>
<td>3</td>
<td>19</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Speakers</td>
<td>550,000,000</td>
<td>230,000,000</td>
<td>104,000,000</td>
<td>90,000,000</td>
<td>80,000,000</td>
<td>66,000,000</td>
<td>80,000,000</td>
<td>47,000,000</td>
<td>46,000,000</td>
<td>37,000,000</td>
<td>31,000,000</td>
<td>13,000,000</td>
<td>4,600,000</td>
<td>50,000</td>
</tr>
<tr>
<td>Editors per Million Speakers</td>
<td>0.1</td>
<td>0.3</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>1</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>3</td>
<td>0.4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
• Let’s simplify this through a Video.......
Generic Challenges

- There are three issues which should be taken into account:
  - a) quality,
  - b) relevance
  - c) consistency of knowledge

- Suitable motivation of the contributors.

- Another issue is the scalability

✉ vishnu@cis-india.org

vishnuvardhan_t
India Specific Challenges for Building Knowledge Platforms on the Internet

• Technical infrastructure for Indian Languages
  – Typing a major challenge
  – OCR – complexity of Indian Language scripts
  – Various other technical troubles like browser compatibility, font display, etc. which deter new users
• Dearth of quality content available in digital format
• Different standards/formats/generations (Gov.in/DLI)
• Relative lack of research/academic standards, which is transferred on to Indic Wikipedias.
• Lack of Knowledge Sharing Culture
• Insularity – Niche Communities
Building a Mass Knowledge Platform

Platform:
• User friendly
• Easily available and adoptable
• Offline outreach is key to effective use of online platforms

Programme:
Feedback loop is key
Behavior statistics data
Reinvent and Replicate the Prg.
Multi-channel awareness
‘User Connect’ programmes

People
• Communicate knowledge sharing objectives
• Make knowledge sharing fun
• Appoint Ambassadors
• Virtual volunteer community building looks simple but complex - Failure

✉ vishnu@cis-india.org

vishnuvardhan_t
Image Sources

- Slide No. 4: 
  http://commons.wikimedia.org/wiki/File:The_Caxton_Celebration_-_William_Caxton_showing_specimens_of_his_printing_to_King_Edward_IV_and_his_Queen.jpg
- Slide No. 5: 
  http://commons.wikimedia.org/wiki/File:Low_Memorial_Library_Columbia_University_NYC.jpg
- Slide No. 6: http://commons.wikimedia.org/wiki/File:Pieni2.jpg
Welcome to accessing and building Knowledge platforms in Digital Era

T. Vishnu Vardhan
Prg. Director Access to Knowledge, Centre for Internet & Society
✉ vishnu@cis-india.org  🚥 vishnuvardhan_t