

Community Efforts in Information Localization of Wikipedia

Presentation for Action Week for Global
Information sharing (AGIS'10)
2010 Dec 7, New Delhi

Arjuna Rao Chavala, Wikipedian
<http://tech4society.blogspot.com>

Email: arjunaraoc@googlemail.com

Copyright: CC-BY-SA-3.0

Agenda

- Wikipedia
 - Site
 - People and Institutions
 - Hardware
 - Software
- Commercial translation tools impact on Community
- Critical Steps for bridging digital divide



వికీపీడియా
స్వేచ్ఛా విజ్ఞాన సర్వస్వము

మొదటి పేజీ | చర్చ | సాధ్యు చూడు | చరితం

Not able to view the Telugu script? [Click Here](#)

తెవికీపీడి మీ అభిప్రాయాలు రాయండి | **Wiki in 5 minutes** | **వరిచయం** | **ప్రయోగశాల** | **టైపింగు సహాయం** | **సహాయం** | **ఈ వారము సమైక్య కృషి** | **రచ్చబండ**

వికీపీడియాకు స్వాగతం!

వికీపీడియా ఎవరైనా రాయదగిన ఒక స్వేచ్ఛా విజ్ఞాన సర్వస్వము. ఇది మామూలు వెబ్ సైట్ల వంటిది కాదు. ఇక్కడ సమాచారాన్ని చూడటమే కాదు, ఉన్న సమాచారంలో అవసరమైన మార్పుచేర్పులు చెయ్యవచ్చు. ఇక్కడ లేని సమాచారాన్ని చేర్చవచ్చు కూడా. ప్రస్తుతం తెలుగు వికీపీడియాలో 41,356 వ్యాసాలున్నాయి. పూర్తి గణాంకాలు చూడండి.

- అకౌంటు వలన లాభాలు
- లాగిన్ పేజీ
- ఎలా తోడ్పడవచ్చు?
- ప్రయోగశాల
- సహాయ కేంద్రం
- రోజుకొక చిట్కా

వరిచయం • అన్వేషణ • కూర్చడం • ప్రశ్నలు • సహాయము

విహారణ • విశేష వ్యాసాలు • అ-ఐ సూచి

స్వాగతం

- వికీపీడియాలో **మీ ఊరు ఉందా?**
- **తెలుగులో వ్రాయడమెలానో** తెలుసుకోండి.
- వికీని త్వరగా అర్థం చేసుకునేందుకు **5 నిమిషాల్లో వికీ** పేజీని చూడండి.
- ఇంకా లోతుగా వెళ్లేముందు వికీపీడియా యొక్క **ఐదు మూలస్థంభాల** గురించి చదవండి.
- వికీపీడియా గురించి తెలుసుకునేందుకు **తరచు అడిగే ప్రశ్నలు** చూడండి.
- సహాయము లేదా **శైలి మాన్యువల్** చూడండి.
- ప్రయోగశాలలో ప్రయోగాలు చెయ్యండి.
- వికీపీడియాకు సంబంధించిన సందేహాలుంటే **సహాయ కేంద్రంలో** అడగండి. మిగిలిన ప్రశ్నలకు **రచ్చబండ** చూడండి.
- చేయవలసిన పనుల గురించి **సముదాయ పందిరిలో** చూడండి.
- వికీపీడియాలో జరుగుతూ ఉన్న మార్పుచేర్పులను చూడాలంటే **ఇటీవలి మార్పులు** చూడండి.

ఈ వారపు వ్యాసము

ఆంధ్ర రాష్ట్ర సాధన కొరకు ఆమరణ నిరాహారదీక్ష చేసి, ప్రాణాలర్పించి, **అమరజీవి** యైన మహాపురుషుడు, **పాట్లీ శ్రీరాములు**, ఆంధ్రులకు ప్రాతఃస్మరణీయుడు. భాషాప్రయుక్త రాష్ట్రాల ఏర్పాటుకు కారణభూతుడైనవాడు. **మహాత్మా గాంధీ** బోధించిన సత్యము, అహింస, హరిజనోద్ధరణ అనే ఆశయాలకొరకు జీవితాంతం కృషిచేసిన మహనీయుడు.



పాట్లీ శ్రీరాములు 1901 మార్చి 16న మధాసు, జార్జిటోను, అణ్ణాపిళ్ళె వీధిలోని 165వ నంబరు ఇంటిలో గురవయ్య, మహాలక్ష్మమ్మ దంపతులకు జన్మించాడు. వారి పూర్వీకులది ప్రస్తుత శ్రీ పాట్లీ శ్రీరాములు నెల్లూరు జిల్లా లోని **పటమటపల్లె** గ్రామం. ఇరవై యేళ్ళు వరకు శ్రీరాములు విద్యాభ్యాసం మధాసు లోనే జరిగింది. తరువాత **బొంబాయిలో** శానిటరీ ఇంజనీరింగు చదివాడు. తరువాత "గ్రేట్ ఇండియన్ పెనిస్యుల్ రైల్వే"లో చేరి దాదాపు నాలుగేళ్ళు అక్కడ ఉద్యోగం చేసాడు. 1928లో వారికి కలిగిన బిడ్డ చనిపోవడం జరిగింది. తరువాత కొద్ది రోజులకే అతని భార్య కూడా చనిపోయింది. 25 యేండ్ల వయసు కలిగిన శ్రీరాములు జీవిత సుఖాలపై విరక్తి చెంది ఉద్యోగానికి రాజీనామా చేసాడు. ఆస్తిపాస్తులను తల్లికి, అన్నదమ్ములకు పంచివెట్టి, గాంధీజీ అనుయాయుగా సబర్బతి ఆశ్రమం చేరాడు. స్వాతంత్ర్యోద్యమంలో పాల్గొన్నాడు. 1930లో **ఉప్పు సత్యాగ్రహంలో** పాల్గొని జైలుశిక్ష అనుభవించాడు. తర్వాత మళ్ళీ 1941-42 సంవత్సరాలలో సత్యాగ్రహాలు, **కిరీట్ ఇండియా**

ఈ వారపు బొమ్మ



- మార్గదర్శకము
- మొదటి పేజీ
 - రచ్చబండ
 - సముదాయ పందిరి
 - వర్తమాన ఘటనలు
 - ఇటీవలి మార్పులు
 - యాదృచ్ఛిక పేజీ
 - కొత్త పేజీలు
 - విరాళములు
- సహాయము
- సహాయసూచిక
 - టైపింగు సహాయం
 - 5 నిమిషాల్లో వికీ
 - వరిచయము
 - ప్రయోగశాల
- వెతుకు
-
-
- పరికరాల పేజ్జీ
- ఇక్కడికి లింకున్న పేజీలు
 - సంబంధిత మార్పులు
 - ప్రత్యేక పేజీలు

- Français
- Polski
- हिन्दी

కొలనుపాక , నల్గొండ జిల్లా, ఆలేరు మండలానికి చెందిన చరిత్రాత్మకమైన గ్రామము. కొద్దికాలం క్రితమే ఒక జైన శివాలయం పునరుద్ధరించబడింది. నలుపు రంగులో శ్రేష్ఠమైన రత్నంతో నిర్మించబడిన మాణిక్య దేవ ఋషభ దేవ విగ్రహము, ఫిరోజా రాతితో నిర్మించబడిన భగవాన్

స్పృహ తప్పి పోయినా దీక్ష నిర్విఘ్నంగా కొనసాగాలని కోరుకొన్నారు. స్పృహ లేనప్పుడు ఎవరూ బలవంతంగా ఇంజెక్షను ద్వారా ఆహారం ఎక్కించరాదని ఆయన చెప్పారు. అప్పటికే రాష్ట్రంలో జరుగుతున్న అలరును హతీరికించాడు.వారివాసం: పాతవి

Telugu Site <http://te.wikipedia.org>

Wikipedia People and Institutions

- Volunteer Contributors build and manage Wikipedia content/tools
- National chapters promote Wikipedia and liason with institutions/stakeholders
- Wikimedia Foundation runs servers

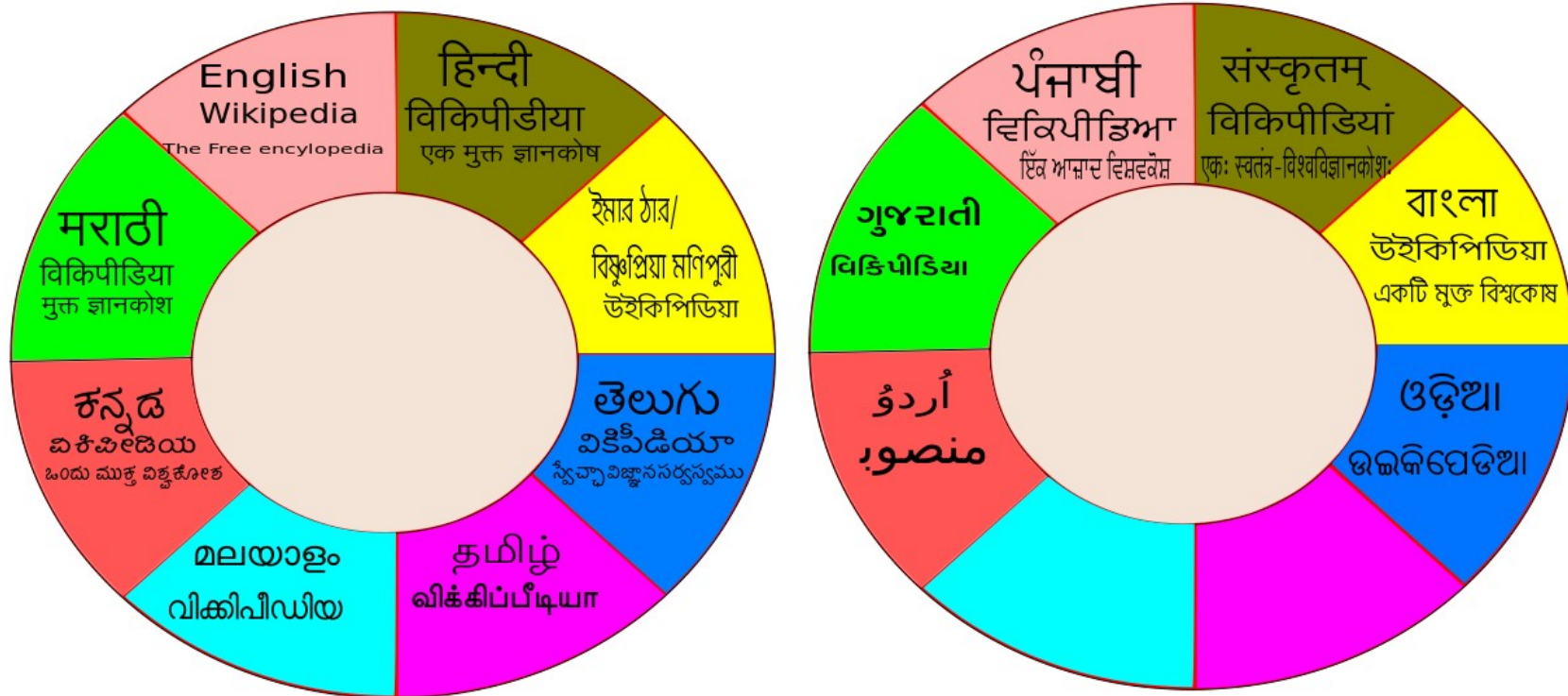
Volunteer Contributors

- What kind of people are these?
- Why do they volunteers?
-
- Let's see couple of short videos
- **Video1, Video2**



Wikipedia

in Indian Languages



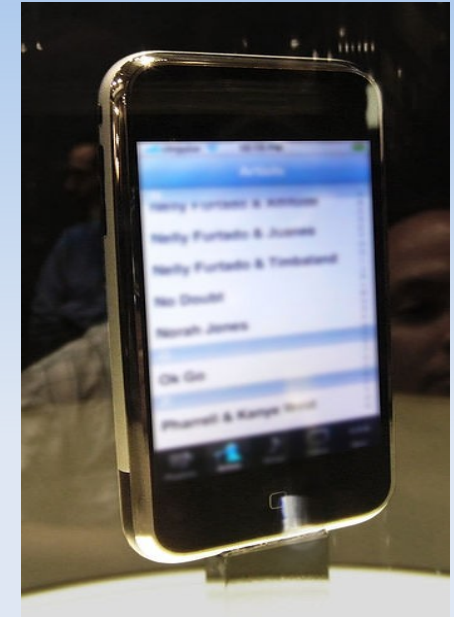
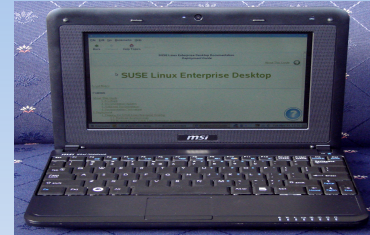
Wikipedia Indian Languages Page View Trend Nov2008-Nov2010

	All languages		Tamil				Telugu								
	English		Hindi		Marathi		Bengali			Malayalam					
	Σ	en	hi	ta	mr	bn	te	ml							
Nov 2009 ⇒ Nov 2010	+21%	7th +21%	9th +16%	12th +14%	14th +6%	19th -13%	23th -55%	20th -41%							
Nov 2010	7,092 M	7,071 M	4.9 M	2.7 M	2.5 M	2.0 M	1.7 M	1.6 M							
..day	236 M /d	236 M /d	162 k /d	91 k /d	82 k /d	65 k /d	55 k /d	53 k /d							
..hour	9.9 M /h	9.8 M /h	6.7 k /h	3.8 k /h	3.4 k /h	2.7 k /h	2.3 k /h	2.2 k /h							
..min	164 k /m	164 k /m	112 /m	63 /m	57 /m	45 /m	38 /m	37 /m							
..sec	2.7 k /s	2.7 k /s	1.9 /s	1.1 /s	1.0 /s										
Top month	Apr 2010 8,240 M	Apr 2010 8,214 M	Sep 2010 5.3 M	Apr 2008 3.2 M	Feb 2010 3.0 M	Feb 2010 2.9 M	Feb 2010 4.5 M	Apr 2010 3.2 M							
Trend last 24 months															
		Kannada		Bishnupriya Manipuri		Gujarati		Nepali							
		Urdu		Nepal Bhasa		Burmese		Sinhala							
		ur	kn	bpy	new	gu	my	si	ne						
Nov 2009 ⇒ Nov 2010	16th -5%	6th +22%	21th -52%	22th -52%	8th +18%	1st +133%	5th +24%	2nd +98%							
Nov 2010	1.1 M	1000 k	649 k	606 k	552 k	373 k	355 k	289 k							
..day	36 k /d	33 k /d	22 k /d	20 k /d	18 k /d	12 k /d	12 k /d	9.6 k /d							
..hour	1.5 k /h	1.4 k /h	902 /h	842 /h	766 /h	519 /h	493 /h	401 /h							
..min	25 /m	23 /m	15 /m	14 /m	13 /m	9 /m	8 /m	7 /m							
..sec															
Top month	Feb 2010 1.2 M	May 2010 1.3 M	May 2008 1.9 M	May 2008 1.6 M	Feb 2010 687 k	Nov 2010 373 k	Apr 2010 460 k	Nov 2010 289 k							
Trend last 24 months															
		Sanskrit		Punjabi		Pali		Assamese		Kashmiri		Sindhi		Oriya	
		sa	pa	bh	pi	as	ks	sd	or						
Nov 2009 ⇒ Nov 2010	17th -6%	4th +47%	3rd +51%	13th +8%	15th +1%	10th +15%	18th -7%	11th +14%							
Nov 2010	191 k	185 k	138 k	93 k	89 k	60 k	54 k	47 k							
..day	6.4 k /d	6.2 k /d	4.6 k /d	3.1 k /d	3.0 k /d	2.0 k /d	1.8 k /d	1.6 k /d							
..hour	265 /h	256 /h	192 /h	129 /h	124 /h	83 /h	74 /h	65 /h							
..min	4 /m	4 /m	3 /m	2.2 /m	2.1 /m	1.4 /m	1.2 /m	1.1 /m							
..sec															
Top month	Apr 2010 277 k	Apr 2010 226 k	Mar 2008 243 k	Mar 2008 179 k	Oct 2010 159 k	Feb 2008 89 k	Apr 2008 193 k	May 2010 59 k							
Trend last 24 months															

Steady growth in Hindi, Tamil, Marathi, Urdu, Kannada, Gujarati, Punjabi

HW Environment

- Desktop/laptop/net-book use growing rapidly
- Cellphones achieved phenomenal growth rate.
 - Every other person in India has a cell phone
- Limited support of Indic language support on cell phones
- Smart phones with support for Indic languages at Consumer price points entering the market (Nokia C3 at about Rs5000(Approx 100USD))



Smart Phones at Consumer price point are critical

SW Environment-Linux

Language	Code	Gnome	Firefox	Mediawiki
भोजपुरी	bh	0.00%	0.00%	22.74%
বাংলা	bn	87.00%	83.00%	88.43%
ইস্ফারঠার/বিস্বপ্রিয়া মনিপুরী	bpy	0.00%	0.00%	51.39%
ગુજરાતી	gu	81.00%	91.00%	39.53%
हिन्दी	hi	80.00%	89.00%	94.74%
ಕನ್ನಡ	kn	81.00%	87.00%	63.41%
मैथिली	mai	50.00%	89.00%	8.80%
മലയാളം	ml	82.00%	83.00%	99.67%
मराठी	mr	83.00%	91.00%	80.81%
नेपाली	ne	53.00%	0.00%	80.07%
नेपालभाषा	new	0.00%	0.00%	12.39%
ଓଡ଼ିଆ	or	80.00%	87.00%	1.43%
ਪੰਜਾਬੀ	pa	92.00%	89.00%	36.06%
தமிழ்	ta	87.00%	82.00%	85.45%
తెలుగు	tcy	0.00%	0.00%	23.39%
తెలుగు	te	82.00%	80.00%	99.96%
اردو	ur	0.00%	0.00%	40.46%

Colorcoding <50%=Red, >=50% and <75%=Yellow and >=75% Green

Bengali, Hindi, Malayalam, Tamil, Telugu -Good, Urdu needs help

Localization Tools

- Offline
 - Poedit
 - Virtaal
- Online
 - Launchpad.net
 - Translatewiki.net

Multiple tools, each with it's own or limited Translation Memories

Translation Tools

- Open source Tools
 - Wiki
 - Translate Toolkit
- Commercial Tools
 - Google Translator toolkit
 - Microsoft WikiBhasha

Simple Open source Translate tools in use by Community.

Commercial Translation Initiatives impact on Community

- Google contributing to Indic language Wikipedia by Paid translators using Google Translator Toolkit
- Engagement is uneven, with some Wikipedias refusing content and some others accepting content with checks
- Process for engaging volunteers is being improved.
- WikiBhasha recently launched. Impact yet to be evaluated.

Critical steps for bridging digital divide

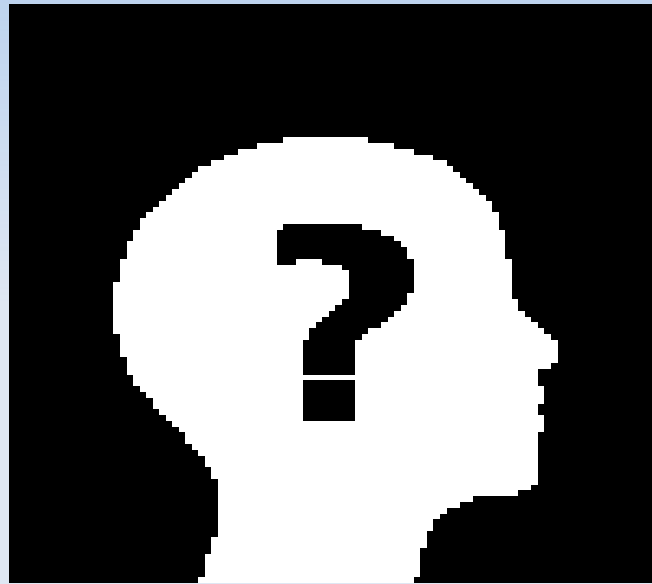
- Commercial contributions of content need to respect the culture and policies of language Wikipedia.
 - Open source engagement of commercial companies like IBM, Sun could be a role model.
- Awareness of native language use on computing devices need to increase. Eg: Wiki Academies
- Information access should become universal.
 - Kiosks and mobile phones reach need to increase.
- Smart phones and data usage charges have to become affordable.
- Government funded tools should be open sourced and put in public domain



Acknowledgements

- Images from Commons
 - Question icon, Netbook, iphone
- Wikipedia Page View data
- Open Translation Tools @ Floss Manuals
- Feedback from Wikipedians Gautam John, Sundar and others on draft version of this presentation

Questions



Email: arjunaraoc@googlemail.com