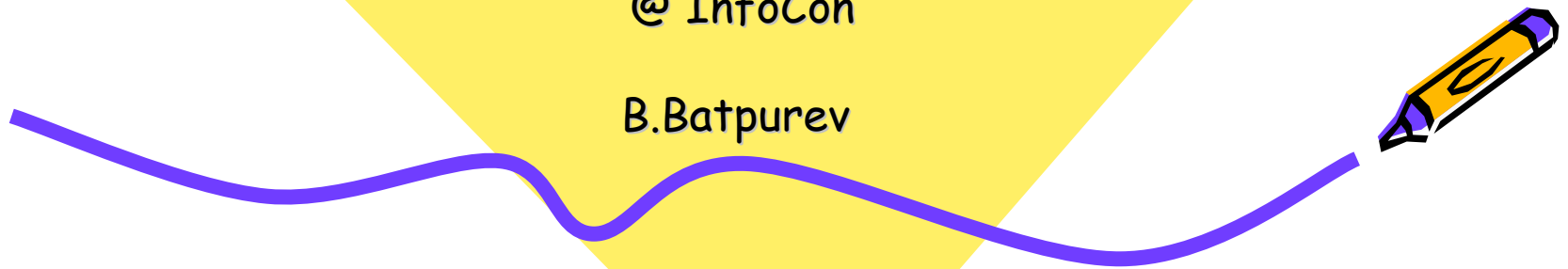


Localization of Open Source tools

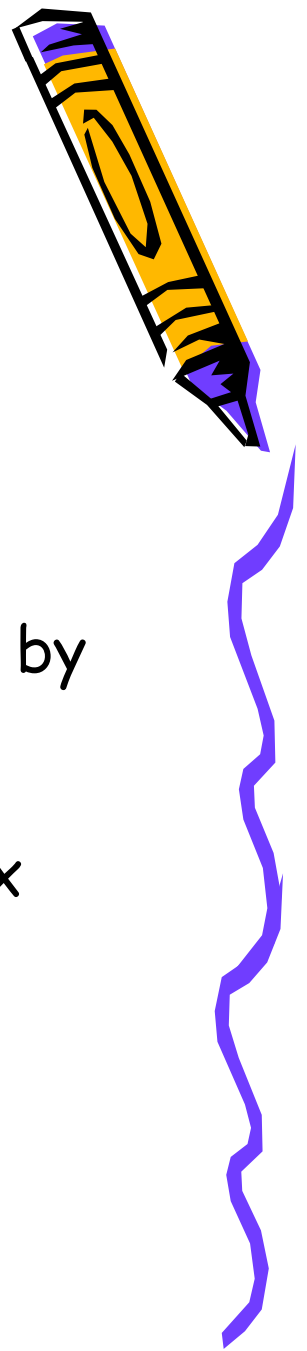
PAN localization phase II
Mongolia component
@ InfoCon

B.Batpurev

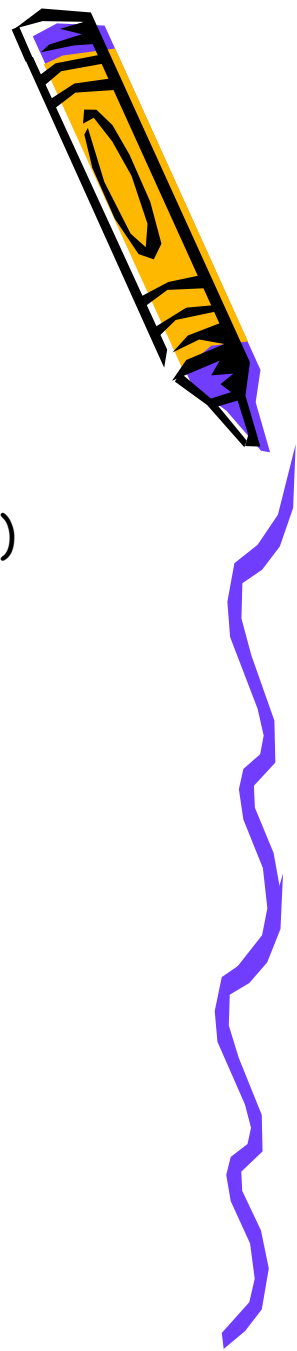


Mongolia component:

- Build a corpus for Mongolian language for computer-based processing.
 - National University of Mongolia
- Development of Mongolian Speech recognition by evaluating and training existing tools
 - Mongolian University of Science and Technology
- Localize OpenOffice, Thunderbird and Firefox
 - InfoCon (country component leader)



Tools already localized

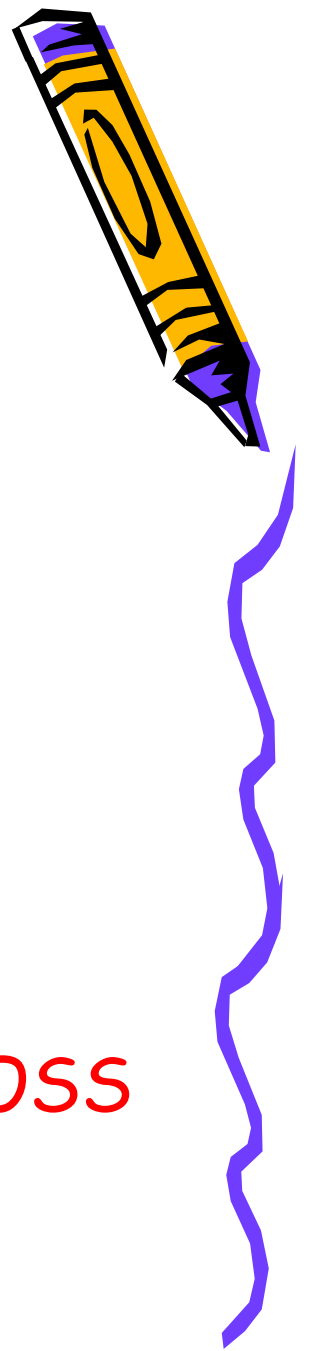


- Advanced tools
 - GFMIS (propriety software)
 - Mongolian TTS (researchers at InfoCon, MUST and NUM)
- Some part of GNOME and KDE is translated
- Content management systems
 - PHP-NUKE (CMS) - (MnDG, InfoCon)
 - DG-Market portal - (MnDG)
 - Greenstone e-library (HSUM)
 - Joomla CMS (ICTA initiative)
 - Moodle LMS - (IDRC)
 - Etc.,



Tools to be localized

@InfoCon



- Firefox ~ 53000 words
- Thunderbird ~ 51000 words
- OpenOffice ~ 153000 words
- GAIM, TLDs ~ 8000 words

Total no of words involved: ~ 265,000

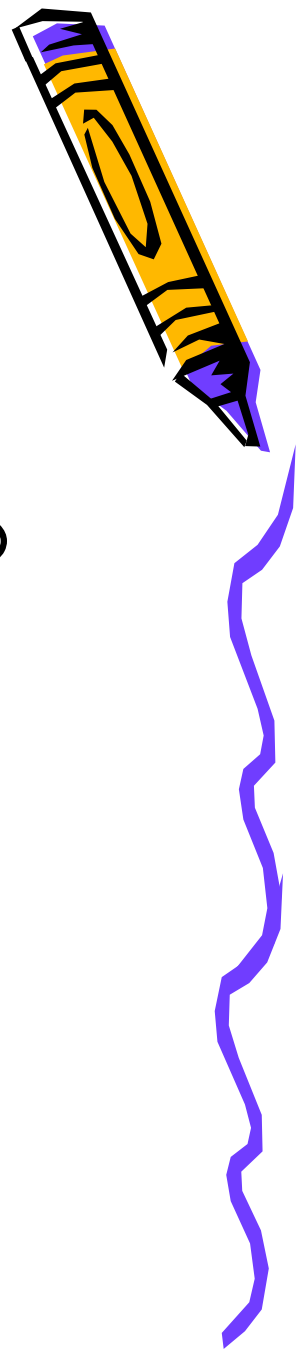
We are going FOSS



OSS utilization

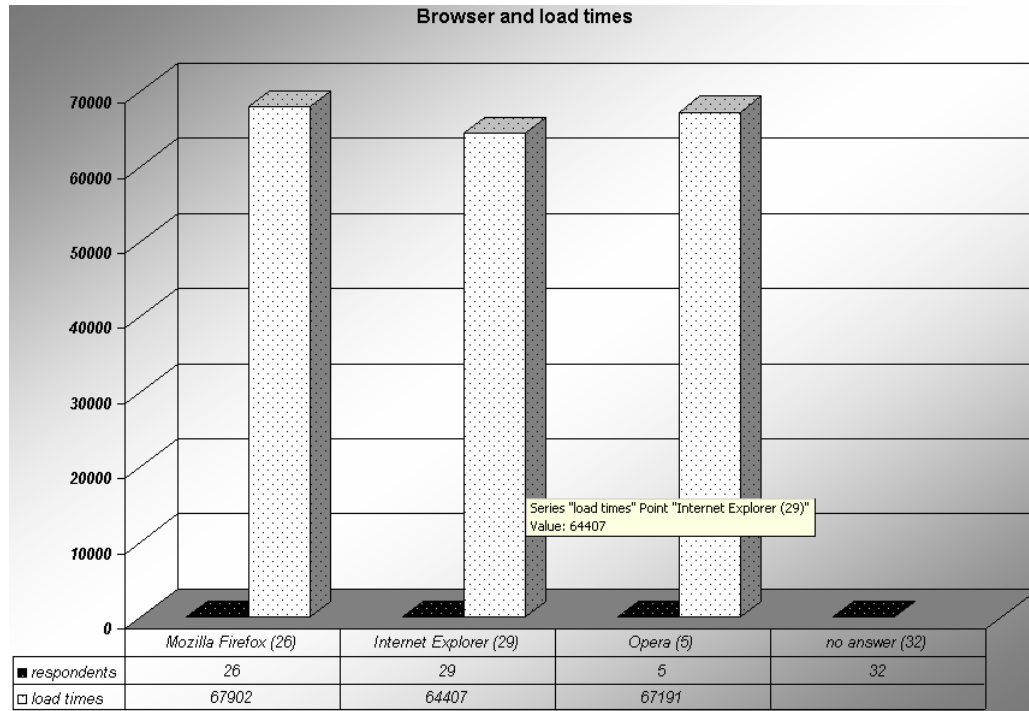
Advanced users: ISPs and Other advanced users

- 9 of 15 ISPs use OSS
 - Server operating system (Redhat, FreeBSD etc.,)
 - Networking (BIND, Samba etc.,)
 - Web serving (Apache, Tomcat)
 - Mail serving (Sendmail, Mailman)
 - Database serving (MySql, PostgreSql)
 - Security solutions (SPAM, AV etc.,)

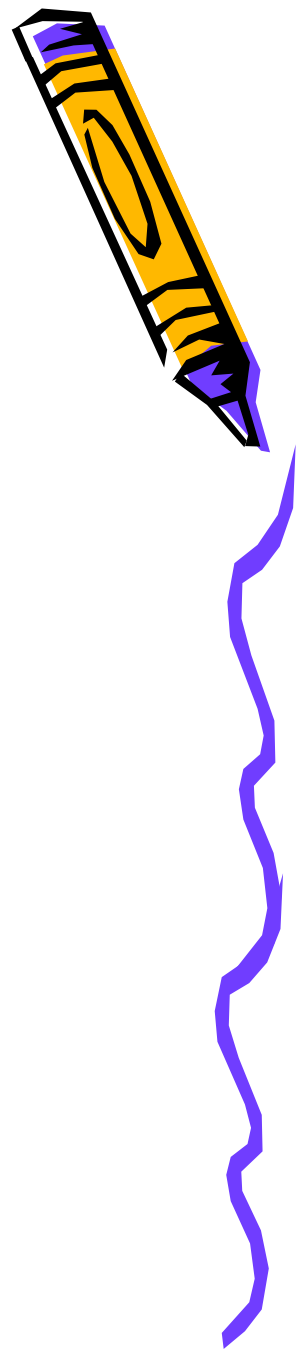


OSS utilization

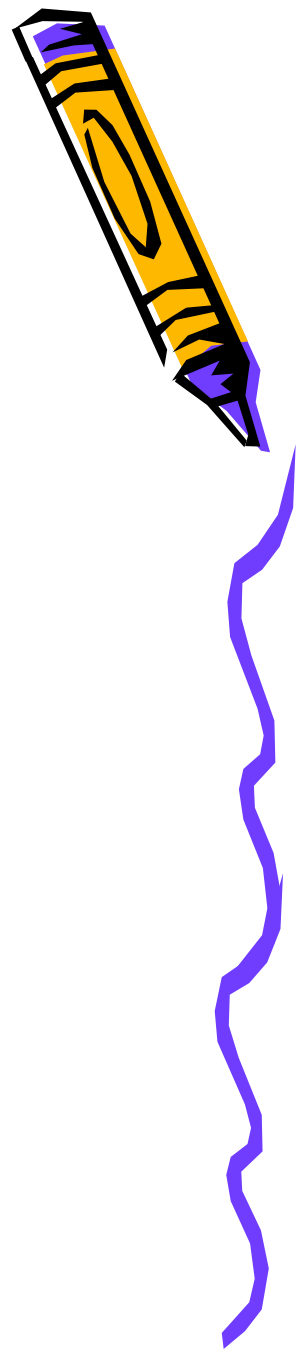
browsers used by end users PANDora survey in Mongolia



- Total respondents - 60
- OSS browser - 31
- Propriety browser - 29



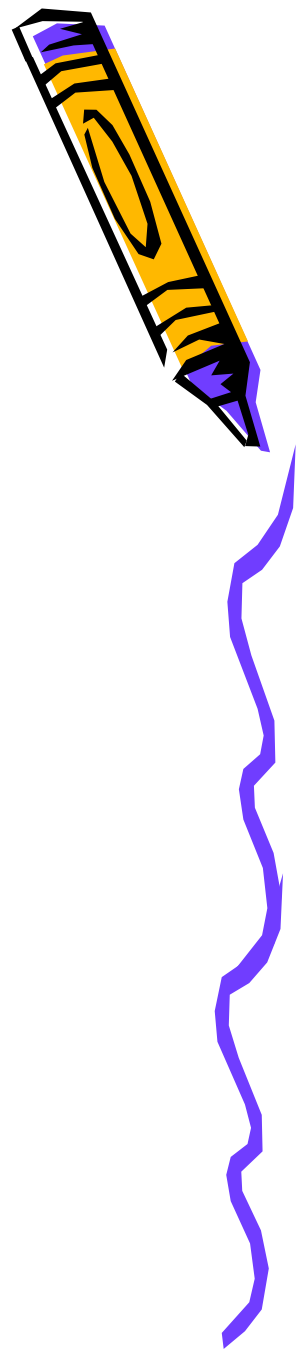
Local online content



- Officially registered entities - 24722 (as of June 2005)
 - 1500 web sites with Mongolian content - 2005
 - 73.6% under ".mn" domain
 - 50.4% owned by private entities
 - 30.7% NGO and Public entities
 - 9.1% govt organizations
 - 9.8% individuals
 - 5.5% of entities have presence on the internet
 - Year 2006 - Mongolian websites over 2000.
- Internet Bandwidth
 - 78Mbs in 2004
 - 247Mbs in 2006
 - Cost reduction 10 times



Timeline



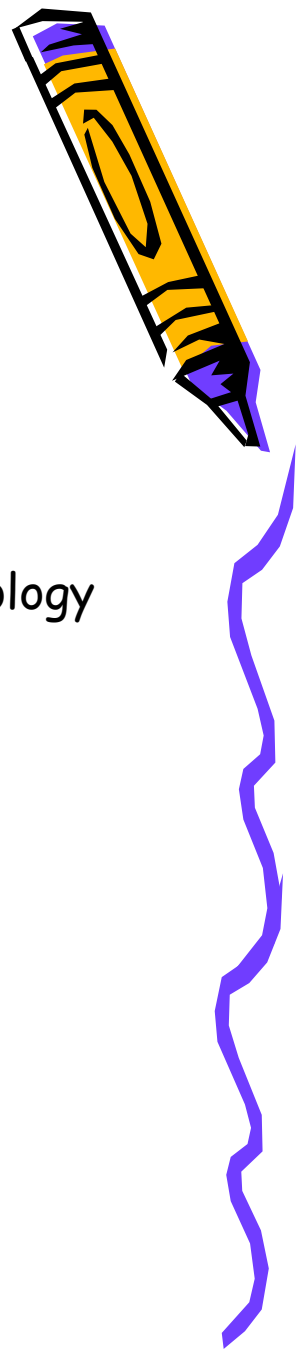
- Initial training, identify tools etc: May - July 2007
- Localization of FF: July 07 - March 08
- Localization of TB: July 07 - May 08
- Open Office Localization: March 08 - May 09
 - Word: March - Sept 2008
 - Spreadsheet: May - Dec 2008
 - Presentation: Sept 08 - March 09
 - Integration and Testing April 2009
 - Release: May 2009

Total duration: May 07 - May 09

OM and GEM Reports submitted at each phase



Highlights



- Translated terms standardization
 - Working group (ICTA, Univ, NGO and Companies)
 - Reference tool (for team and public)
 - Mongolian National Center for Standardization and Metrology (MNCSM)
- Develop and Strengthen
 - NLP, LLC - Awareness, Human resource and Networking
 - LLC policy
- Wider utilization of FOSS,
 - IPR issues (prevent software piracy etc.,)
 - Reduce costs associated for utilization of ICTs
 - Etc.,



IT terms dictionary

15000 words www.itdic.edu.mn



Үг	Орчуулга	Түвшин	Ангилал
(Broadband Metropolitan area network) BMAN	(Батлаагүй)	Ерөнхий	Техник хангамж (сүл)
(CIR) committed information rate	(Батлаагүй)	Ерөнхий	Техник хангамж (сүл)
(Complementary Code Keying) CCK	(Батлаагүй)	Ерөнхий	Техник хангамж (сүл)
(High definition)	(Батлаагүй)	Ерөнхий	Хадгалах байгууламж
(High speed)	(Батлаагүй)	Ерөнхий	Хадгалах байгууламж
(INCLUSIVE-OR) element	(Батлаагүй)	Ерөнхий	Ерөнхий үгс(сүл)
(INCLUSIVE-OR) gate	(Батлаагүй)	Ерөнхий	Ерөнхий үгс(сүл)
(complete) connection in telecommunication	(Батлаагүй)	Ерөнхий	Ерөнхий үгс(сүл)
(digital) transmission path	(Батлаагүй)	Ерөнхий	Ерөнхий үгс(сүл)
(electrical) circuit	(Батлаагүй)	Ерөнхий	Ерөнхий үгс(сүл)
(keyboard, video, mouse)	(Батлаагүй)	Ерөнхий	Хадгалах байгууламж
(layer) service	(Батлаагүй)	Ерөнхий	Ерөнхий үгс(сүл)
(mean time between failures)	(Батлаагүй)	Ерөнхий	Хадгалах байгууламж

MNS (8000terms)

Section 2: + Terms and definition

01 Fundamental terms 01.01 General terms

01.01.01 Information (in information processing)

Knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts that within a certain context has a particular meaning.

01.01.02 data

A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.

NOTE:
Data can be processed by humans or by automatic means.

Хоёрдугаар бүлэг: Нэр томьёо ба тодорхойлолт

01 Үндсэн нэр томьёо 01.01 Ерөнхий нэр томьёо

01.01.01 мэдээлэл (мэдээлэл боловсруулахад)

Баримт, үйл явдал, эд зүйлс, процесс зэрэг объект ба утга бүхий тодорхой агуулгын хүрээнд авч болох ойлголтыг багтаасан санааны талаархи мэдлэг.

01.01.02 өгөгдөл

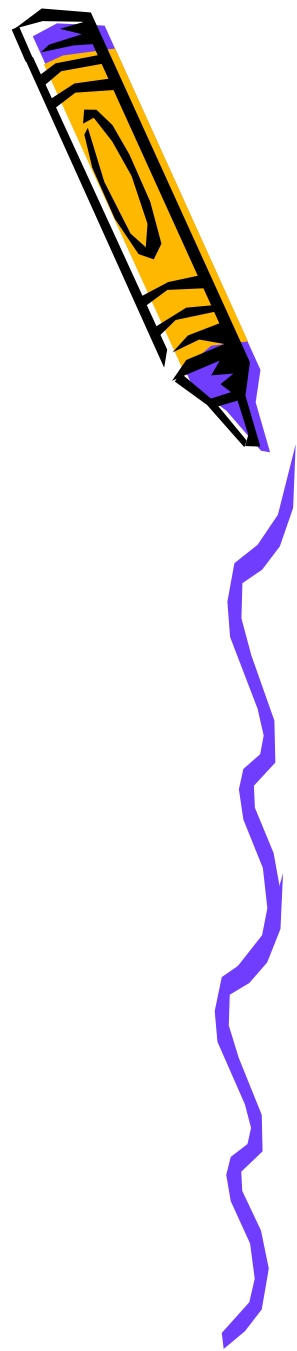
Мэдээллийг дамжуулах, тайлах, боловсруулахад зохимжтой хэлбэрээр дахин тайлах дүрслэл.

ТАЙЛБАР:
Өгөгдлийг хүн буюу автомат хэрэгслээр боловсруулж болно.



Outputs

- OpenOffice in Mongolian
- FF, TB and GAIM in Mongolian
- Standard translation of involved terms



Thank you
&
Tashi Delek

B.Batpurev, InfoCon
batpurev@infocon.mn

