

**8th World Telecommunication/ICT Indicators
Meeting (WTIM-10)**
Geneva, Switzerland, 24 - 26 November 2010



Contribution to WTIM-10 session 7

Document C/32-E
25 November 2010

English

SOURCE: Funredes/Mayaa

TITLE: Indicators for linguistic diversity on the Internet: the way forward (WSIS Target 9)




PARTNERSHIP ON
MEASURING ICT
FOR DEVELOPMENT

Indicators for linguistic diversity on the Internet: the way forward (WSIS Target 9)

Target 9

Encourage
the development of **content**
and put in place technical conditions
in order to facilitate
the presence and use
of all world **languages**
on the Internet

Daniel Pimienta
pimienta@funredes.org
**NETWORKS & DEVELOPMENT
FOUNDATION**



<http://funredes.org>

<http://funredes.org/LC>

**MEMBER OF EXECUTIVE
COMMITTEE**
<http://maaya.org>



maaya
Red Mundial para la Diversidad Lingüística

Presentación de la Red Maaya
Autoridades y Estructura de la Red
Miembros
Actividades
Reuniones
Novedades
Estatutos
Cómo ser miembro
Documentos para descargar

Maaya – Red Mundial para la Diversidad Lingüística – es una red multilateral creada con el propósito de contribuir a la valorización y promoción de la diversidad lingüística en el mundo.
Su sede está ubicada en Ginebra, Suiza.
► [Presentación de la Red Maaya](#)

Target 9 indicators today

- ✓ Proportion of Internet users by language
- ✓ Proportion of webpages per language
- ✓ Number of domains names per Country Code Top Level Domain (under discussion)

FEW MONTHS AGO...

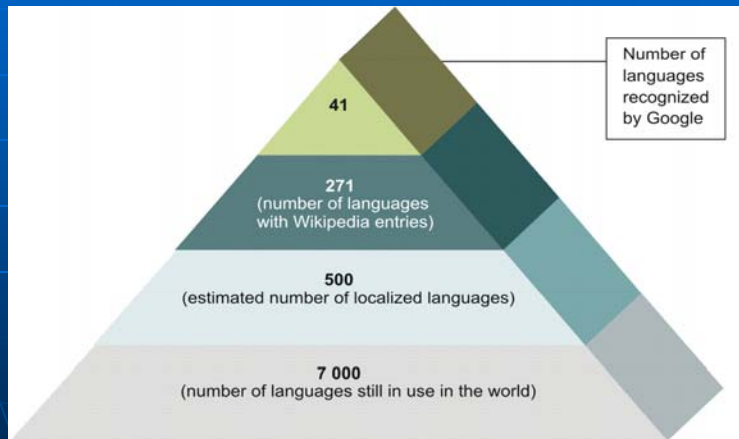


WSIS Forum 2010

"Measuring the WSIS Targets"

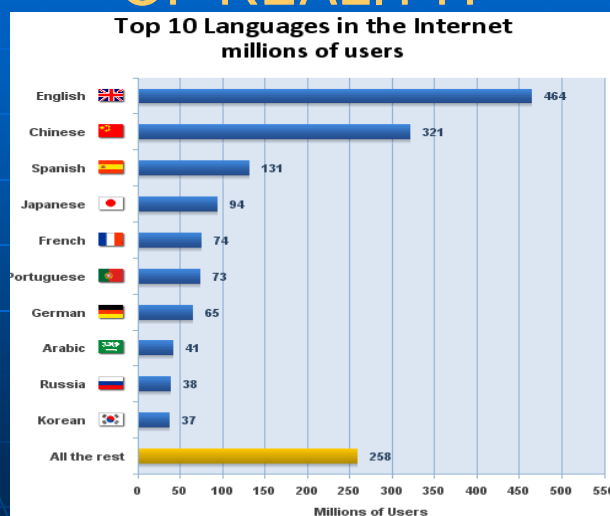
10 May 2010
Geneva

LETS GET A ROUGH FIGURE OF REALITY



Sources : Ethnologue (2010), Wikipedia (2010) and Google (2010).

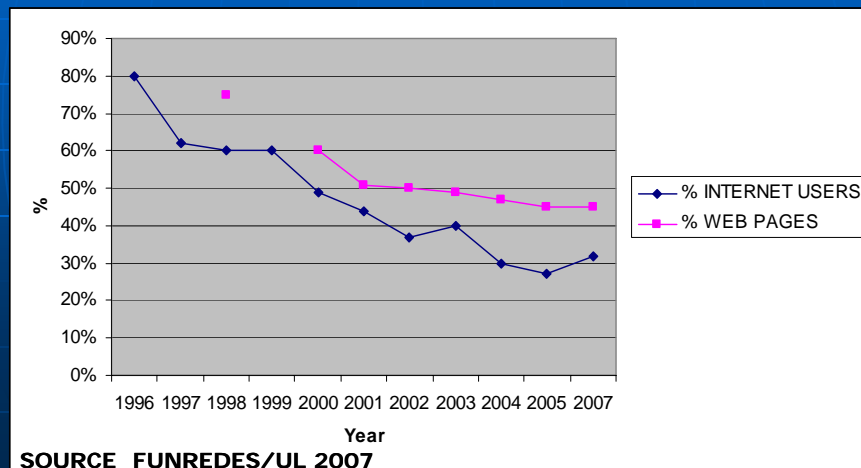
LETS GET A ROUGH FIGURE OF REALITY!



Source: Internet World Stats - www.internetworldstats.com/stats7.htm
 Estimated internet users are 1,596,270,108 for 2009Q1
 Copyright © 2009, Miniwatts Marketing Group

LETS GET A ROUGH FIGURE OF REALITY

ENGLISH



The Digital Divide is much deeper
in terms of

CONTENTS

than in terms of

ACCESS

AFRICA FOR EXAMPLE

- 4.8 % of Internet users
(source InternetWorldStats – 2010)
- 0.6 % of English web pages
(source FUNREDES/UL 2007)
- 0.6 % of French web pages
(source FUNREDES/UL 2007)
- African local languages each make up between
0,006 % & 0,06 % of total web pages
(source LOP 2007)

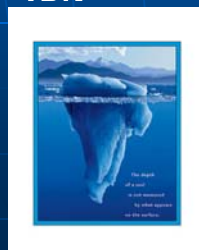
LINGUISTIC DIVERSITY IN THE INTERNET

→ Paradoxical situation

Is the Internet really for every one?

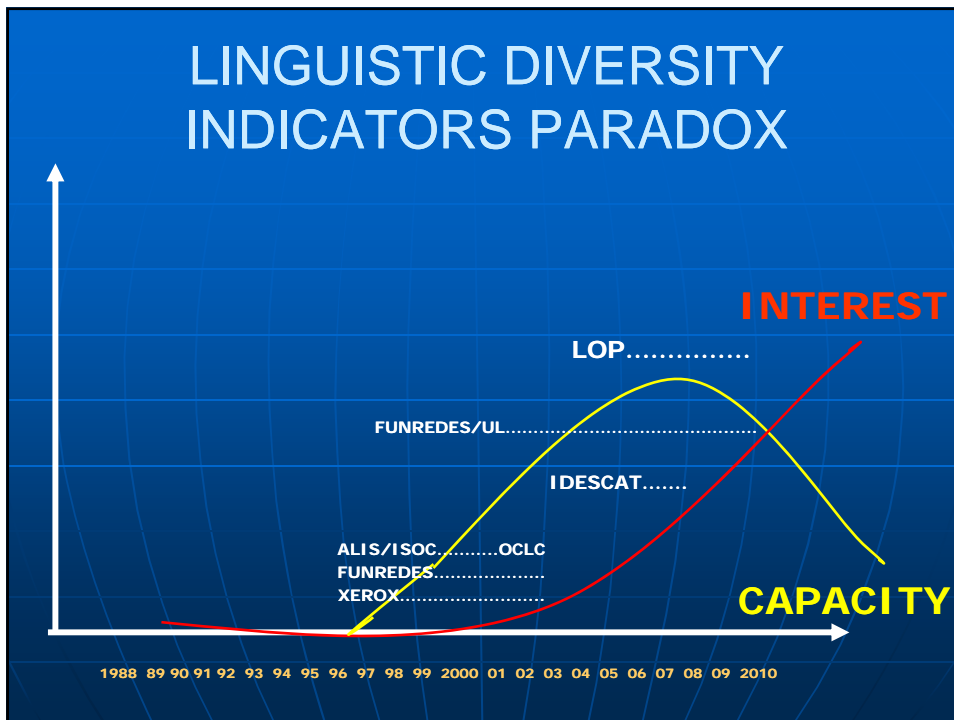
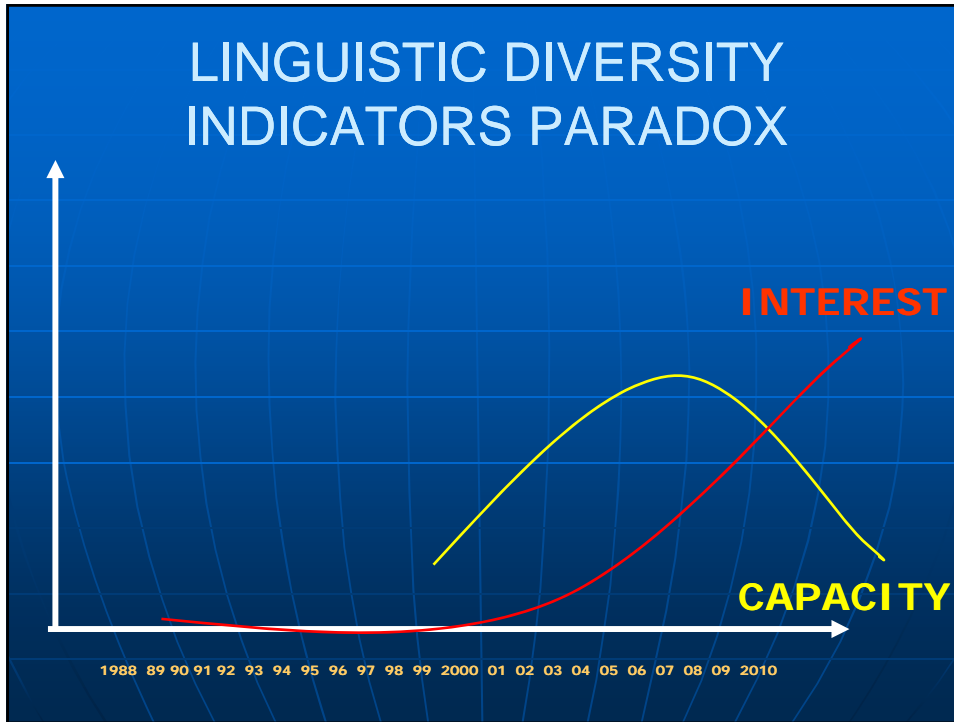


IDN



INDICATORS





WHAT INDICATORS DO WE HAVE?

Internet users per language (source InternetworldStats)

Web pages per language (not all!)

Other indicators per country (FUNREDES/UL)

WHERE IS THE BOTTLENECK?

The two main methods to collect indicators rely on:

- **crawling** TLD domains for languages in Asia and Africa and applying recognition algorithms (LOP)
- using **search engines counting** capacity and their **large percentage of web coverage** (FUNREDES/UNION LATINA)

WHERE IS THE BOTTLENECK?

- The size of the web is getting too large for traditional crawling (**close to infinite!**)
- Search engines are no longer indexing a substantial part of it (**80% → 20%**)
- Search engines' counting data has become **unreliable**
- Data are so far limited to **static data** focusing on the **supply** side. How about trying to understand the **demand** side?

WHAT SOLUTION?

- **Applied research for creating linguistic diversity indicators is **urgently** required**
- **An ambitious and **collaborative** effort with immediate results is needed**

DILINET

A project to construct indicators of linguistic diversity in cyberspace

- **Collaboration between UNESCO, OIF & UNION LATINA with participation of ITU & MAAVA**
- **Project and consortium definition**
- **Targeting to launch an EU/PF7 three years comprehensive project in first semester 2011**

DILINET LINES OF RESEARCH

- **Crossing crawling time limitations by using distributed computing (BOINC) or super-computers (GENSI/PRACE)**
- **Overcoming crawling sequentiality by using mathematical tricks (e.g. cell automatas or statistical methods)**
- **Using new statistical approaches derived from opinion surveys**
- **Focusing the demand side (offering user behaviour watch programs like in alexa.com)**
- **Opening new avenues with other formats (like audio file language recognition)**

DILINET EXPECTED PRODUCTS

- New and complete line of indicators for linguistic diversity in cyberspace with a concern for **sustainability** and frequency for data production
- Capacity to **crosscheck and assess** results leading to an increase in confidence
- Enlarging the perspective with **new formats** and surveyed areas
- Sensitization and capacity building for policy makers to integrate linguistic diversity in **virtual policies**

DILINET PARTNERSHIP

COMMITTED: UNESCO, OIF, UNION LATINA, FUNREDES, MENON

PROBABLE : ITU, LOP

REQUIRED : European Universities or Research Centers with following skills: mathematics, statistics, computing (crawling), language processing (voice)

DILINET PARTNERSHIP

We are still looking for partners

If you are interested to join,
contact :

pimienta@funredes.org

If you are interested in further
information, check (later):

<http://dilinet.org>

REFERENCES (1/3)

"Twelve years of measuring linguistic diversity in the Internet: balance and perspectives", Unesco/2009.

[http://portal.unesco.org/ci/en/ev.php-URL_ID=29594
&URL_DO=DO_TOPIC&URL_SECTION=201.html](http://portal.unesco.org/ci/en/ev.php-URL_ID=29594&URL_DO=DO_TOPIC&URL_SECTION=201.html)

REFERENCES (2/3)

- "Accessing Contents", D. Pimienta, Chapter of Global Information Society Watch, APC, HIVOS, ITEM, 2008
<http://www.giswatch.org/gisw2008/thematic/AccessingContent.html>

- Measuring linguistic diversity on the Internet, UNESCO, 12/2005, a collection of papers by: J. Paolillo, D. Pimienta, D. Prado, et al.
http://portal.unesco.org/ci/en/ev.php-URL_ID=20882&URL_DO=DO_TOPIC&URL_SECTION=201.html

REFERENCES (3/3)

- "Quel espace reste-t-il dans l'Internet, hors la langue anglaise et la culture "made in USA" ?", in « Nord et Sud numériques », Les Cahiers du Numériques, Vol 2 No 3/4 Hermès, Numéro spécial sur la fracture numérique, 2001

<http://funredes.org/lc/l5/cahiersNumFinal.html>

- Activités de Funredes pour la promotion de la diversité linguistique dans l'Internet et enseignements de l'expérience, 5/05

http://portal.unesco.org/ci/en/file_download.php/92a27500bf11f4c73cd567943deb4077Daniel+Pimienta.doc

