Comparing Accesses to ETDs and Journals in Education and Languages Available from the Same Repository

Ana Pavani

Member IEEE

Laboratório de Automação de Museus, Bibliotecas Digitais e Arquivos

Departamento de Engenharia Elétrica
Pontifícia Universidade Católica do Rio de Janeiro
Brazil

apavani@lambda.ele.puc-rio.br
http://www.maxwell.lambda.ele.puc-rio.br/

This work is a continuation of two works presented in previous ETD symposia. All three are devoted to analysis of accesses to ETDs available from the Maxwell System (http://www.maxwell.lambda.ele.puc-rio.br), PUC-Rio's repository. But there are significant differences:

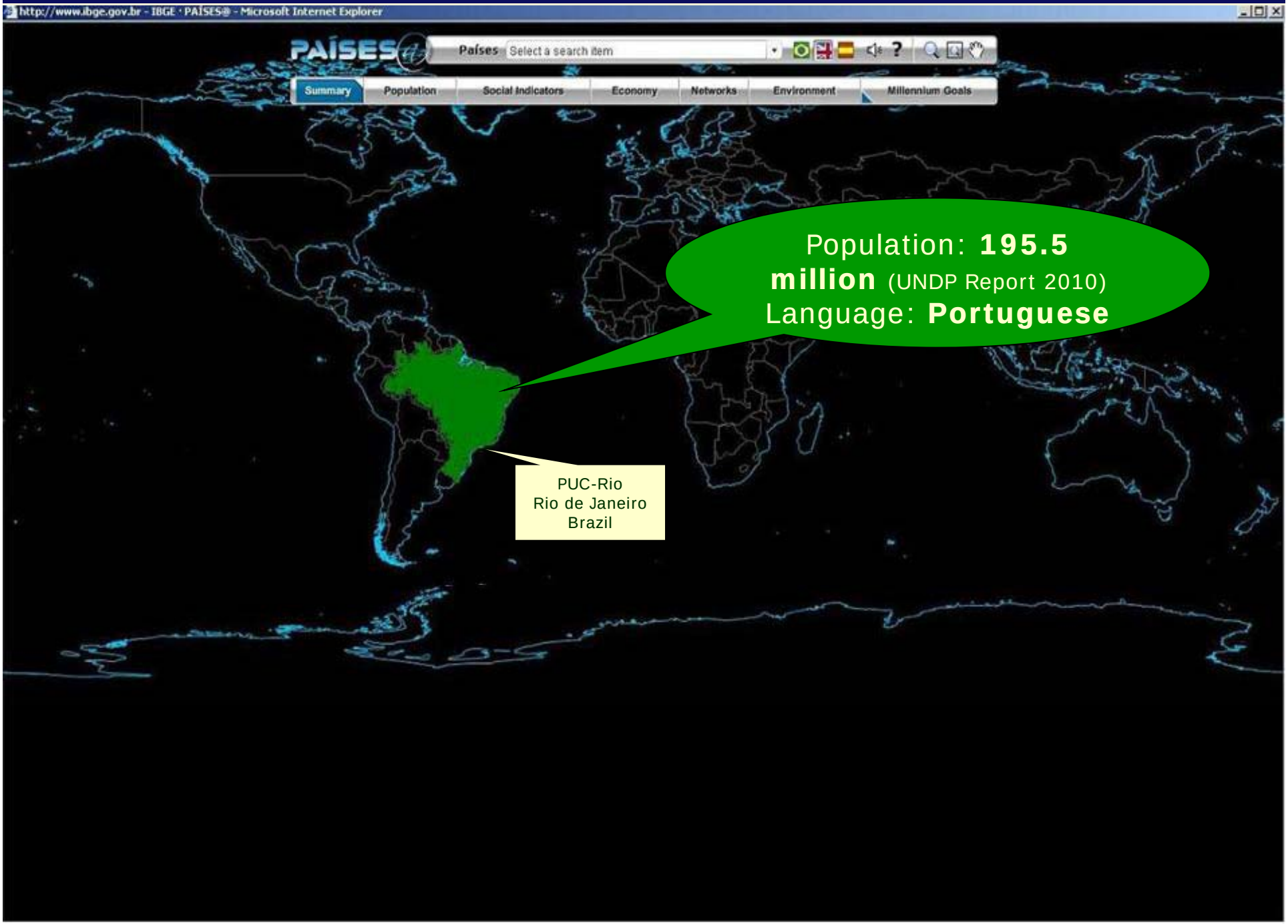
- **This work compares accesses to ETDs and accesses to journals in the same areas of knowledge – the previous works considered accesses to ETDs only**

- This work addresses two areas of knowledge in the Humanities – the previous works took into consideration ETDs in Humanities & Theology, Social Sciences and Science & Technology

- The time frame of this work is from Mar 2010 to Jun 2012 – the previous works had longer time frames

ETDs, JOURNALS, PUC-Rio, DOAJ, BDTD & NDLTD


http://www.ibge.gov.br - IBGE · PAÍSES@ - Microsoft Internet Explorer
PAÍSES@
Países
Select a search item
Summary
Population
Social Indicators
Economy
Networks
Environment
Millennium Goals


http://www.ibge.gov.br - IBGE - PAÍSES@ - Microsoft Internet Explorer
PAÍSES@
Países
Select a search item
Summary
Population
Social Indicators
Economy
Networks
Environment
Millennium Goals
Population: 195.5 million (UNDP Report 2010)
Language: Portuguese
PUC-Rio
Rio de Janeiro
Brazil

PUC-Rio is a small private university. It is divided in 3 centers and each has graduate programs:

- CTCH (Humanities) – 6
- CCS (Social Sciences) – 10
- CTC (Science & Technology) – 12

- The oldest graduate program (EE) started in 1963.
- The newest graduate program is less than 5 years old.

Characteristics of PUC-Rio's ETD program:

- First published ETD – May 2000
- ETDs became mandatory – Aug 2002
- Number of ETDs – 6,191 (Jun 2012)
  - CTCH – 1,454
  - CCS – 1,465
  - CTC – 3,272
- Yearly average number of defended T&Ds(*) – ≅520

- (*) 2007, 2008, 2009, 2010 & 2011;
- There is retrospective digitization.

	# Mar 10	% Mar 10	# Jun 12	% Jun 12
Humanities & Theology	1,239	24.52	1,454	23.49
Social Sciences	1,072	21.22	1,465	23.66
Science & Technology	2,741	54.26	3,272	52.85
All	5,052	100.00	6,191	100.00

PUC-Rio's ETDs, BDTD(*) and NDLTD (**) :

- Number of BDTD institutions – 96 (OAI-PMH data providers)
- Number of BDTD metadata records – 200K+ (BDTD is an OAI-PMH data and service provider)
- BDTD records are/were harvested by many institutions, and made available worldwide
- Brazilian ETDs are the largest collection in Portuguese available worldwide

- (*) BDTD – Biblioteca Digital de Teses e Dissertações = Brazilian Nat'l Consortium.
- (**) You must know what NDLTD stands for!!!

Characteristics of PUC-Rio's journals:

- First published journal – Jun 2003
- Number of journals – 11 (Jun 2012)
  - CTCH – 9 (3 in Education, 4 in Languages, 1 in Theology and 1 in Design)
  - CCS – 2
- Journals indexed on DOAJ – Directory of Open Access Journals (http://www.doaj.org/) – 9

Accesses to PUC-Rio's digital collection:

- Access logs saved since – Jun 2004
- Number of monthly logs in Sep 2012 – 100

OPTIONS IN THIS WORK

Options in this work are:

- Areas of knowledge – Education and Languages due to the number of journals in these areas
- Journals – 3 in Education and 3 in Languages (the 4th journal in Languages was not considered because the first issue was published in Jun 2012)
- Time frame – Mar 2010 to Jun 2012 due to the fact that Mar 2010 was the month in which the newest journal considered in this analysis started being published
- Computations – all performed in terms of % since numbers are different in the two areas

ETDs:

	Mar 10	Jun 12	Average
Education (*)	186	229	208.64
Languages (*)	421	486	466.04

- (*) 100% in Portuguese.

Journals:

		Mar 10	Jun 12	Average
Education (*)	# issues	13	21	16.86
	# items	148	223	181.39
Languages (**)	# issues	19	28	23.79
	# items	165	387	328.75

- (*) Jun 2012 – 2.01% in en, 97.32% in pt and 0.67% in es.
- (**) Jun 2012 – 2.34% in en, 95.55% in pt and 2.11% in es.
- These % will change because items in foreign languages have been allowed in the last couple of years.

pt & es IN THE WORLD

	Worldwide	Western Languages	Internet
Portuguese	7th	3rd	6th
Spanish	2nd	1St	3rd

pt is the official or one of the official languages of:

- Angola
- Brazil
- Cape Verde
- Equatorial Guinea (*)
- East Timor (**)
- Guinea-Bissau
- Macau (***)
- Mozambique
- Portugal
- Sao Tome and Principe

es is the official or one of the official languages of:

- Argentina
- Bolivia
- Chile
- Colombia
- Costa Rica
- Cuba
- Dominican Rep
- Ecuador
- El Salvador
- Equatorial Guinea (*)
- Guatemala
- Honduras
- Mexico
- Nicaragua
- Panama
- Paraguay
- Peru
- Puerto Rico
- Spain
- Uruguay
- Venezuela

(*) es & pt official
(**) less than 5% of the population know it; it was banned during the Indonesian rule
(***) UNDP did not publish in the last report; other data were used

Assumptions for the analysis:

- ETDs and journals are very specialized items – people who seek them are highly educated
- es and pt are quite similar languages – educated people who can speak one can read the other
- es and pt-speakers are potential readers of PUC-Rio's ETDs and journals
- 2 countries were considered separately:
  - Brazil – is the home country
  - US – there are very large groups of es and pt-speaking persons but neither one is the language of the country

- 3 other groups were defined:
  - “pt group” – all countries that have pt as one of the official languages
  - “es group” – all countries that have es as one of the official languages
  - “international group = others” – all countries except Brazil, the US and pt and es-speaking countries
- Factors considered to influence accesses to scholarly works:
  - Population size
  - Level of education
  - Access to the Internet

DEALING WITH COUNTRIES DIFFERENCES


Mexico has 110M inhabitants
Sao Tome and Principe has 165K inhabitants
Portugal and Spain are in Europe
Angola and Mozambique are in Africa
Argentina and Honduras are in Latin America
Portugal has 10M inhabitants
Spain has 45M inhabitants
Equatorial Guinea has the 2 languages

Quantization of potential accesses from countries that are very different :

- Need to find data on the factors that may influence accesses to ETDs:
  - Population size – easy
  - Level of education – difficult (literacy rates are easy!)
  - Access to the Internet – difficult
  - All data should be considered in the same time frame
- Knowledge that the second and the third factors are dependent on how developed countries are
- Knowlede that it was necessary to combine the 3 factors

Decision on how to deal the countries differences:

- Use UNDP's HDI – Human Development Index that contains information on the second and the third factors (HDI combines indicators of life expectancy, education and income; the new way it is computed contains means years of schooling and expected years of schooling, going beyond literacy rates)

- Decision to combine HDI with the population size

Index I = Population x HDI

	es-speakimg	pt-speaking
Total population	420,281,000	57,858,800
Average HDI	0.707	0.527
Index I	**309,420,871**	**25,114,111**

Comments:

- 21 es-speaking and 10 pt-speaking countries
  (Equatorial Guinea was counted in both)
- Average HDI for es-speaking countries is 34.16% higher than the other group
- Population of the es-speaking countries is almost 7.4 times the population of the other group
- Index I for the es-speaking group is 12.36 times the same index for the pt-speaking group

The expectation was to have many more accesses from es-speaking countries than from pt-speaking countries!!

WORKING WITH DATA AND RESULTS

Information:

- Number of sets of data – 28 (one for each month)
- For each set, 16 variables were computed (examples – number of countries, number of pt-speaking countries countries, total number of accesses, etc)
- All data were computed for Education (ETDs and journals) and Languages (ETDs and journals)
- For both ETDs and journals there are monthly production (publishing) numbers and accesses numbers

Access numbers:

- The Apache Server log is processed to identiffy rows that indicate accesses to contents; this happens once every hour. Identification means: (1) the content ID on the system; and (2) the country (from an IP x country international table) where accesses came from

- The result of the processing is included on a table of the database that stores numbers of accesses to each accessed content from each country in a year-month. This means that this table constains a row for each content-country-year-month and the last datum of the row is the corresponding number of accesses

This procedure yields data that combined with Production Statistics are the source of the results that known as **Access Statistics**.

Access Statistics qualify when and where accesses came from as well as the sets that the accessed items belong to. An example of set is journal and issue, a second example is graduate program and level.

The main results in the paper follow

ETDs - % accesses:

	Education	Languages	Complete Collection
Brazil	82.01	79.37	78.47
United States	10.71	8.23	11.85
Pt-speaking group	3.44	5.67	4.78
Es-speaking group	0.39	0.68	0.73
Others	3.44	6.06	4.18

Journals - % accesses:

	Education	Languages	Complete Collection
Brazil	86.48	86.90	86.76
United States	2.22	4.19	3.54
Pt-speaking group	4.59	3.01	3.53
Es-speaking group	2.96	1.01	1.66
Others	3.75	4.89	4.51

Obvious comments:

- In all cases accesses from Brazil show the highest numbers!
- In all cases of journals accesses from Brazil are higher than the ones corresponding to ETDs, though ETDs in the areas are 100% in pt while journals items are not
- Accesses from the US are higher in the case of ETDs than in the journals
- Accesses from the es-speaking group are higher in the case of journals than in the case of ETDs

- Accesses from “Others” are not significantly different in both cases
- In the case of ETDs, Languages is “more international” than Education, but both are below the complete collection
- In the case of ETDs, accesses from the US are approximately 10% and more significant than the other groups

Other comments:

- The Maxwell Systems is an OAI-PMH data provider with 2 separate URL – one for ETDs (http://www.maxwell.lambda.ele.puc-rio.br/ibict.php) and one for all contents (http://www.maxwell.lambda.ele.puc-rio.br/DC_Todos.php)
- ETD metadata go to BDTD, NDLTD, Scirus, Scientific Commons, UNDLP, UNM, etc
- Journals are only indexed on DAOJ
- Google indexes ETDs and journals

THANK YOU! ¡MUCHAS GRACIAS! OBRIGADA!