International Visibility of ETDs in English and in Portuguese on a Brazilian Repository

Ana M B Pavani, *Member IEEE*

Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, Brazil

Abstract

Electronic Theses and Dissertations (ETDs) are very important outputs of the educational process at Graduate Level. Authors, supervisors, Graduate Programs and Institutions benefit when ETDs are visible worldwide. One of the challenges of international visibility is related to the languages of the works. English is the *lingua franca* of the Internet as well as of scholarly communication. Institutions in countries where English is not one of the official languages may have different policies towards ETDs publication. This work addresses the visibility of two different sets of ETDs (in Portuguese and in English) on the repository of an institution in a country whose official language is Portuguese but works can be published in other languages (they are not translations). Accesses to the two sets are examined from data collected from the repository that hosts them; the time frame is 2019-2023. Compilation and examination of data showed that the percentages of accesses from Portuguese speaking countries to ETDs in English are significantly smaller than the corresponding numbers in Portuguese, as expected. The most important result is that the percentages of accesses from countries that have neither Portuguese nor English as official languages is almost four time as big when ETDs in English are compared to the ones in Portuguese. Tables and graphics are presented to show the results.

Keywords: ETDs in English, ETDs in Portuguese, international visibility, accesses

Background and Introduction

This section presents the context of this work. It is divided in three subsections. The first is devoted to the visibility / accessibility of ETDs, the second shows some data on the Portuguese language in the world. Finally, the third addresses some characteristics of ETDs at Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).

Visibility / Accessibility of ETDs

ETDs are important for the students, mentors, graduate programs and universities. They are rich sources of up to date information that can be used as references for research – they are useful to graduate students developing theses or dissertations. They also showcase the quality of research at the institutions. International visibility indicates the interest on the topics addressed in the works and may result in increased references to them. Visibility of ETDs is a concern all over the world and expressed with different terms. At the same time, university administrators and managers of ETD Programs are concerned with visibility from different points of view.

Paskali et al. (2021) addressed the subject as "discoverability" and described a very comprehensive work at the University of Novi Sad, Serbia, that took care of all the steps of the PhD Dissertations lifecycle.

An interesting solution also came from Serbia. Ivanović and Ivanović (2016) when overcoming the problem posed by 90% of the Doctoral Dissertations defended in the country being published in the Serbian language. The Serbian Ministry of Science has prescribed a rule that all dissertations must have, at least, one paper published in a journal belonging to the Web of Science JCR list prior to its defense. The CRIS – Current Research Information System of the Novi Sad University links both, dissertation and article, thus bringing visibility to the dissertaion.

A different approach to visibility was addressed by Weideman (2015). The author analyzed (a very limited) sample of ETDs from India available on the National Repository – Shodhganga

(https://shodhganga.inflibnet.ac.in/). The author's concern was with the way Google Scholar (GS) indexed the works both in term of the metadata and the full text.

Abdelrahman and Fox (2022) focused on automatically extracting and classifying information from ETDs to generate metadata from abstracts, titles and keywords and, then classify by subject based on the metadata. The challenge comes from both the complexity of the language and the lack of previous research. The goal of the research was to improve accessibility to Arabic ETDs.

Another concern about visibility of ETDs is their accessibility to readers that are visually impaired or blind.

Over 15 years ago Pavani (2008) presented an ETDs accessibility project that was under development at PUC-Rio; it had been motivated by a suggestion from students with vision problems. The project was completed and it results implemented on the Maxwell System (http://www.maxwell.vrac.puc-rio.br) which is the Institutional Repository (Lynch, 2003) of PUC-Rio; instructions to authors on how to generate accessible files for their ETDs, unfortunately, were not addressed since they required additional efforts.

This topic is so important that Ingram et al. (2023) foused on it with a wider scope of special needs. They stress the difficulties in accessing PDF (Portable Document Format) files and propose a solution solution based on AI techniques to support the conversion of PDF files to HTML versions.

Finally, another aspect of visibility was addressed in the context of Open Access Elecctronic Theses and Dissertations by Ramirez et al. (2013). The work discusses the eventual diminishing of publication of the contents of ETDs in Social Sciences and Humanities if they are in Open Access.

The works mentioned in this section address the enhancement of accessibility / visibility with actions to take in the ETD creation process, including the decision on the level of access they will have.

Portuguese in the World

Portuguese is the 9th most spoken language (4th Western language) in the world according to Berlitz (https://www.berlitz.com/blog/most-spoken-languages-world) or the 8th (4th Western according to Statista (https://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide/). Both numbers mean that Portuguese has a large number of speakers. The Comunidade de Países de Língua Portuguesa (https://www.cplp.org/) lists nine countries that have Portuguese as one of the official languages – Angola (36.7M), Brazil (216.4M), Cape Verde (0.6M), Guinea-Bissau (2.2M), Equatorial Guinea (1.7M), Mozambique (33.9M), Portugal (10.2M), Sao Tome & Principe (0.2M) and Timor-Leste (1.4M). Their populations in 2023 were informed on WorloMeter (https://www.worldometers.info/world-population/population-by-country/).

It can be observed that Brazil holds 71.3% of the Portuguese speaking population in the world. This has consequences in the results of this work.

ETDs at Pontifícia Universidade Católica do Rio de Janeiro

Pontifícia Universidade Católica do Rio de Janeiro has offered graduate level programs since the early 1960s, when theses and dissertations were published in Portuguese unless a special request was submitted (and approved) to use a different language. The ETD program started in 2000 and ETDs became mandatory in 2002; in 2008, the use of foreign languages stopped requiring a special authorization. The current (September 27, 2024) number of ETDs is 15,342 and the numbers is by language are:

Portuguese (pt) – 14,006 (91.29%); English (en) – 1,301 (8.49%); Spanish (es) – 30 (0.2%); French (fr) – 2 (insignificant); Italian (it) – 1 (insignificant)

No doubt Portuguese dominates the collection at the moment though the numbers of ETDs in English have monotonically been growing. The number of ETDs in Spanish is surprising low since PUC-Rio has had a large number of students from Spanish speaking countries of Latin Ameica for many decades.

There are Graduate Programs in three different areas: (1) Science & Technology; (2) Social Sciences; and (3) Theology & Humanities. The first Graduate Programs were in Science & Technology. Currently, all Programs have the Masters and Doctoral levels, except for Architecture and Metrology. The percentages of ETDs in English are higher in Science & Technology and Economics when compared to other areas of knowledge.

ETDs are made available from the Maxwell System. It hosts many more items of scholarly communication and also Educational Resources. All items, except Educational Resources, have DOIs – Digital Object identiifiers. In order to make it easier to find ETDs, since the number of items on the IR is large, an aggregator was deployed – ETDS @PUC-Rio (https://www.maxwell.vrac.puc-rio.br/projetosEspeciais/ETDs/).

Objective and Methodology of this Work

Differently from the works referenced in the previous section, this paper does not address suggestions to make ETDs more visible / accessible. It concerns the analysis of accesses of ETDS published in two languages. Its objective is to map accesses from different parts of the world to two sets of the ETDs on the repository. The sets are the ones with the highest numbers of ETDs – Portuguese and English. It is important to remark that the ETDs are not translations from Portuguese to English and vice-versa; they were originally written in each language. The time frame of the observation is set of five years 2019 - 2023.

The characteristics of accesses to be observed are: (1) the numbers of countries where the accesses came from; (2) the distribution of accesses among Brazil, pt-speaking countries; en-speaking countries and the United States. Data of accesses are available from dataset Accesses to Electronic Theses and Dissertations in English and in Portuguese (Pavani, 2024). The choice for the last five full years is due to: (1) the numbers of ETDs in English has become more significant; and (2) the accesses were counted over full years. In order to see the numbers of ETDs in different languages over the years:

- Access link https://www.maxwell.vrac.puc-rio.br/est_serie_por_ppg_ano_lingua.php;

- On the center top of the page switch the language to "English";

- Choose PUC-RIO – ALL in the combo box;

- Wait a little bit, it may take some time.

The data in the return table is updated in real time every hour. It is easy to see that the number of ETDs in English is getting higher as time goes by. For this reason, it is important to observe the characteristtics of the publishing and of the accesses. Then the Research Question can be formulated.

The Research Question is:

> *Are the patterns of accesses (numbers of countries, languages of the countries and numbers of accesses) different for ETDs in English and ETDs in Portuguese?*

This work is based on the analysis of data gathered from the Institutional Repository that are organized in different ways so that conclusions are reached.

Analysis of Data in the Dataset

Initially, an overview is presented and the data on the accesses are discussed.

An Overview of the Data

Data in the dataset were consolidated in table 1 to yield an overview of the numbers. The information is analyzed from different view points.

Table 1

Overview of data on profiles of ETDs and accesses to them (in the dataset) from 2019 to 2023

		2019	2020	2021	2022	2023
1	Number of ETDs in Portuguese	10,828	11,584	12,572	13,110	13,705
2	Average numbers of partitions in Portuguese	5.5	5.3	5.0	4.8	4.7
3	Number of ETDs in English	565	741	968	1,099	1,222
4	Average numbers of partitions in English	2.9	2.5	2.3	2.2	2.0
5	Total ETDs	11,408	12,343	13,570	14,240	14,959
6	Total accesses to ETDs in Portuguese	7,124,454	7,757,880	6,904,436	6,604,265	9,343,634
7	Total accesses to ETDs in English	59,422	62,000	79,527	90,711	129,802
8	Average accesses to ETDs in Portuguese	657.97	669.71	549.19	503.76	681.77
9	Average accesses to ETDs in English	105.17	83.67	82.16	82.54	106.22
10	Average accesses to ETDs in Portuguese by partition	119.63	126.36	109.84	104.95	145.06
11	Average accesses to ETDs in English by partition	36.27	33.47	35.72	37.52	53.11
12	% ETDs in Portuguese	94.92	93.85	92.65	92.06	91.62
13	% ETDs in English	4.95	6.00	7.13	7.72	8.17

Obs: For many years the library required that ETDs be submitted in one file per chapter – this yielded partitioned ETDs. After the requirement was cancelled the average number of partitions has monotonically been decreasing.

- Evolution of the numbers of ETDS in English and in Portuguese

  - The last two rows of table 1 show that the percentages of ETDs in Portuguese and in English. It is clear that the percentages of ETDs in Portuguese are decreasing while the corresponding numbers for English are increasing. This can clearly be observed by examining the Time Series of Published ETDs by Language available at https://www.maxwell.vrac.puc-rio.br/est_serie_por_ppg_ano_lingua.php?nrseqtip=1&ppg=TODPUC; switch the language to English at the center top of the page.

  - Rows 2 and 4 show that the average numbers of partitions of ETDs in English are less than half of the corresponding numbers for ETDs in Portuguese, except in 2019 when the ratio was 57.72%. These numbers are decreasing. For many years, the library required that ETDs be submitted in one file per chapter; this yielded partintioned ETDs. After this requirement was dropped, the average numbers of partitions have monotonically been decreasing.

- Evolution of the average accesses to ETDs in English and in Portuguese

  · Rows 8 and 9 show that the average numbers of accesses to ETDs in English are much smaller than to ETDs in Portuguese.

  · Rows 10 and 11, on the other hand, show that the differences are much smaller when partitions are taken into consideration; ETDs in Portuguese have higher numbers of files and thus accessing an average ETD in Portuguese requires accessing more files than an average ETD in English; average numbers of partitions are shown in lines 2 and 4.

Table 2 shows the numbers of countries that accessed ETDs; these data are from the dataset too.

Table 2

Numbers of countries that accessed ETDs from 2019 to 2023

		2019	2020	2021	2022	2023
1	Number of countries that accessed ETDs	218	227	200	202	204
2	Number of countries that accessed ETDs in pt	215	227	198	194	201
3	Number of countries that accessed ETDs in en	170	164	151	170	152

Obs: The numbers in row 1 contain accesses to ETDs in the other three languages.

The numbers is table 2 came as a surprise. Pavani and Barbosa (2024) studied accesses to Simulator OER - Open Educational Resources available from the Maxwell System. Ten pairs of Portuguese originals and corresponding English versions had accesses tracked during six years. The results showed that the average numbers of countries of origin in the time frame for each pair were higher for the English simulators in six out of the 10 pairs (60%). The results in table 2 show that ETDs in Portuguese were accessed from higher numbers of countries in all five yeas of observation. A possibility for such result is the numbers of ETDs in each language – there ratio between number of ETDs in Potuguese and number of ETDs in English ranges from 19.16 (2019) to 11.22 (2023). A significantly larger number of ETDs offers a lot more topics to choose from.

Analysis of the Accesses

This subsection is devoted to the analysis of accesses to ETDs in terms of the languages of the countries of origin. It does not take into consideration the years when the accesses occurred – they are aggregated in the interval 2019 to 2023. To obtain data, the dataset was considered again – this time separately examining the two sheets for ETDs in English and in Portuguese.

In order to group countries according to the languages, the criteria were:

- Identify as English or Portuguese speaking each country that has the corresponding language as the official or one of the official languages.

- Consider Brazil and the United States as separate from the language groups due to the large populations and numbers of accesses from them.

Compilation of data from the dataset yielded tables 3 and 4, and figures 1 and 2.

Table 3

Numbers of accesses to ETDs in English by language group

	# Accesses	% Accesses
Brazil	64,478	15.30
pt speaking countries	1,750	0.42
US	165,226	39.20
en speaking countries	35,581	8.44
Other countries	154,427	36.64

Table 4

Numbers of accesses to ETDs in Portuguese by language group

	# Accesses	% Accesses
Brazil	28,010,450	74.23
pt speaking countries	2,482,901	6.58
US	3,516,951	9.32
en speaking countries	299,158	0.79
Other countries	3,425,209	9.08

In tables 1, 2, 3 and 4, it is important to observe that the columns # Acceses are quite different with respect to numbers they contain. Some facts must be considered:
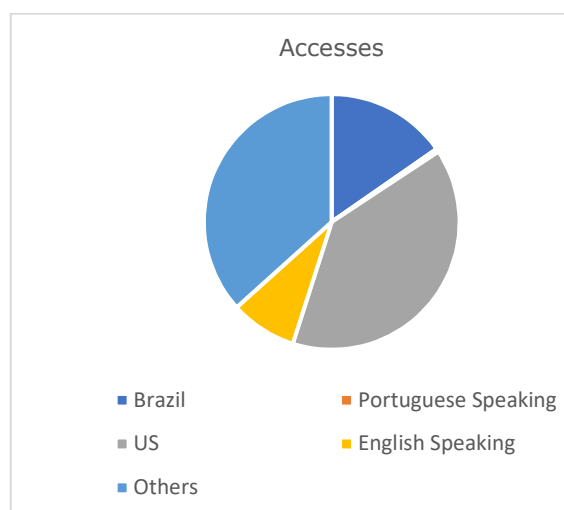
- The average numbers of partitions of ETDs in English are smaller than the corresponding numbers of ETDs in Portuguese for all the years in the time interval.

- ETDs in English are not translations of originals in Portuguese and are mainly in Science, Technology and Economics. Maybe accesses may have been influenced by the topics of the ETDs.

Due to the differences in the numbers of ETDs and in the numbers of partitions, the columns % of Accesses may contains more significant information:
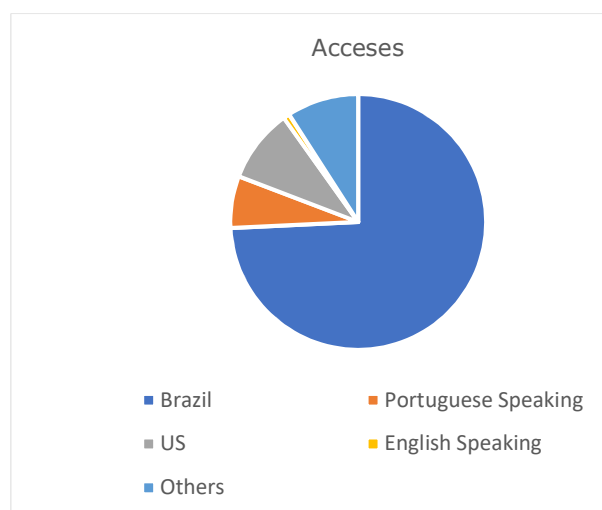
- Accesses to ETDs in English are 39.20% from the United States and 36.64% from countries that do not speak Portuguese nor English – this indicates they are more international. The percentges differ aproximately 10% from each other. Fifure 1 shows this.

- Accesses to ETDs in Portuguese are almost ¾ from Brazil. This is very easy to see in figure 2.

- The percentage of accesses to ETDs in Portuguese originating from the United States or countries that do not speak English nor Portuguese are very similar and smaller that 10%.

- Accesses to ETDs in Portuguese from Portuguese speaking countries is very low. It is important to recall that Brazil has 71.3% of the Portuese speaking population in the world.

- Accesses to ETDs in Portuguese from English speaking countries and to ETDs in English from Portuguese speaking countries are very low. In both cases they account for less than 10%.

Figure 1

Percentages of accesses from different language groups – ETDs in English

Figure 2

Percentages of accesses from different language groups – ETDs in Portuguese

Conclusion and Future Work

The analysis and understanding of the impact of languages of the ETDs on the accesses they have is very important because it can yield information to Graduate Programs to address this issue with students. The trend to write ETDs in English is solid and the percentages have been growing every year; this happens mainly with ETDs in Science & Technology and Economics.

Even with lower numbers, ETDs in English have accesses that are more international. ETDs in Portuguese have the main sources of accesses in Brazil

Current analysis of percentages of ETDs in English across the spectrum of Graduate Programs has shown very different results; this work must be continued in order to identify the patterns in different areas of knowledge. A second topic to address is how accesses to ETDs in English and in Portuguese are impacted by the topics they study and the language they are written.

References

Abdelrahman, E., Fox, E. (2022). "Improving Accessibility to Arabic ETDs Using Automatic Classification".

In: Silvello, G., *et al.* Linking Theory and Practice of Digital Libraries. TPDL 2022. Lecture Notes in

Computer Science, vol 13541. Springer, Cham. https://doi.org/10.1007/978-3-031-16802-4_18

Ingram, W. A., Wu, J. and Fox, E. A. (2023) Maximizing Equitable Reach and Accessibility of ETDs, *2023*

ACM/IEEE Joint Conference on Digital Libraries (JCDL), Santa Fe, NM, USA, pp. 256-257, doi:

10.1109/JCDL57899.2023.00049.

Ivanović, L.; Ivanović, D. (2016) Non-English PhD dissertations and associated English journal

papers. 19th International Symposium on Electronic Theses and Dissertations (ETD 2016): "Data

and Dissertations", Université de Lille Sciences humaines et sociales, Jul 2016, Villeneuve d'Ascq,

France. hal-01398456, available: https://hal.univ-lille.fr/hal-01398456

Lynch, C. (2003), Institutional Repositories: essential infrastructure for scholarship in the digital age,

ARL Bimonthly Report, nº 226, USA, 2003.

Paskali, L.; Ivanović, L.; Ivanović, D.; and Surla, D. (2021) "Six Steps Toward Improving

Discoverability of Ph.D. Dissertations," The Journal of Electronic Theses and Dissertations: Vol. 1 ,

Article 2. DOI: http://doi.org/10.52407/FHAM2773 Available at:

https://scholarworks.uaeu.ac.ae/j-etd/vol1/iss1/2

Pavani, A. M. B. (2008), "Making ETDs Accessible to the Visually Impaired and the Blind: a Project Under

Way", ETD 2008 – 11th International on Electronic Theses and Dissertations, Scotland, 2006.

 Available at http://docs.ndltd.org/metadata/etd2008/13/ and https://www.maxwell.vrac.puc-

rio.br/colecao.php?strSecao=resultado&nrSeq=15648@2

Pavani, A. M. B. (2024), "Accesses to Electronic Theses and Dissertation in English and in Portuguese",

Dataset, Maxwell Repository, Brazil. https://doi.org/10.17771/PUCRio.ResearchData.66494.

Pavani, A. M.B. and Barbosa, W. S. (2024), "Examinig Accesses to Open Access Simulator Objects in English

and Potuguese, presented at ELDL 2024 - 8th e-Learning and Digital Learning, Hungary, 2024,

preprint   available   at   https://www.maxwell.vrac.puc-

rio.br/colecao.php?strSecao=resultado&nrSeq=68171@2

Ramirez, M. L.; Dalton, J. T.; McMillan, G.; Read, M.; and Seamans, N. H.. (2013). Do Open Access

Electronic Theses and Dissertations Diminish Publishing Opportunities in the Social Sciences and

Humanities? Findings from a 2011 Survey of Academic Publishers. College & Research Libraries,

74 (4), 368-380. Available at https://scholar.uwindsor.ca/leddylibrarypub/24.

Weideman, M. (2015), ETD Visibility: A Study on the Exposure of Indian ETDs to the Google Scholar

Crawler, 18th International Symposium on Electronic Theses and Dissertations (ETD 2015):

evolving genre of ETD's for knowledge discovery. India. Available at

https://www.researchgate.net/publication/321868313_ETD_Visibility_A_Study_on_the_Exposu

re_of_Indian_ETDs_to_the_Google_Scholar_Crawler_0168

.