

**SEU SUPER NOME AQUI**

**A MALDIÇÃO DOS TÍTULOS DE TRÊS LINHAS EM  
MONOGRAFIAS DE GRADUAÇÃO: PANORAMA  
HISTÓRICO, CRÍTICAS E PROPOSTAS**

**CIDADE**

**2012**

**SEU SUPER NOME AQUI**

**A MALDIÇÃO DOS TÍTULOS DE TRÊS LINHAS EM  
MONOGRAFIAS DE GRADUAÇÃO: PANORAMA  
HISTÓRICO, CRÍTICAS E PROPOSTAS**

Monografia apresentada à disciplina Orientação Monográfica XYZ como requisito parcial para obtenção do título de Bacharel em XYZ, com ênfase em XYZ, do Setor de XYZ, Universidade XYZ.

Orientador: Nome do seu orientador

**CIDADE**

**2012**

# Termo de Aprovação

SEU SUPER NOME AQUI

## A MALDIÇÃO DOS TÍTULOS DE TRÊS LINHAS EM MONOGRAFIAS DE GRADUAÇÃO: PANORAMA HISTÓRICO, CRÍTICAS E PROPOSTAS

Monografia aprovada como requisito parcial para obtenção do título de Bacharel em XYZ, com ênfase em XYZ, do Setor de XYZ, Universidade XYZ, pela seguinte banca examinadora:

---

Prof. Dr. Fulano  
Universidade XYZ

---

Prof. Dr. Beltrano  
Universidade XYZ

Cidade, 2012

# Dedicatória

A quem eu dedico,

Eu, mim e moi.

# Agradecimentos

Primeira linha dos agradecimentos.

Se precisar, uma segunda linha para mais agradecimentos, afinal você vaza gratidão e quer bajular o maior número possível de professores.

## Epígrafe

— What’s your name?

— Fuck you! That’s my name.

Alec Baldwin, em “Sucesso a Qualquer Preço”

# Sumário

<b>Lista de Figuras.....</b>	<b>vii</b>
<b>Lista de Tabelas .....</b>	<b>viii</b>
<b>Lista de Siglas .....</b>	<b>ix</b>
<b>Lista de Símbolos .....</b>	<b>x</b>
<b>Resumo.....</b>	<b>xi</b>
<b>Abstract .....</b>	<b>xii</b>
<b>1 Organização.....</b>	<b>1</b>
<b>2 Introdução.....</b>	<b>2</b>
2.1 Motivação .....	2
2.2 Objetivos.....	3
<b>3 Fundamentação teórica.....</b>	<b>4</b>
3.1 Seção .....	4
3.1.1 Sub-seção.....	4
<b>4 Metodologia .....</b>	<b>5</b>
4.1 Corpora .....	5
4.1.1 Especificação.....	5
<b>5 Análises .....</b>	<b>6</b>

5.1	Stopwords .....	6
<b>6</b>	<b>Conclusão.....</b>	<b>7</b>
	<b>Referências Bibliográficas .....</b>	<b>8</b>
	<b>Apêndice A – Sugestão de léxico para estudo.....</b>	<b>9</b>
	<b>Apêndice B – Códigos .....</b>	<b>10</b>
B.1	Código em Python externo, embutido no PDF.....	10
	<b>Anexo A – Primeiro anexo.....</b>	<b>12</b>



## Lista de Figuras

Figura 5.1	Frutas em um gráfico .....	6
Figura A.1	<a href="http://publicdomainreview.org/2012/07/09/arabic-machine-manuscript">http://publicdomainreview.org/2012/07/09/arabic-machine-manuscript</a>	12

## Lista de Tabelas

Tabela A.1	Lemas mais usados por Cícero (cima pra baixo, esquerda pra direita)	. 9
------------	---	-----

## Lista de Siglas

NLTK	Natural Language Toolkit
LC	Linguística de Corpus
API	Application Programming Interface

## Lista de Símbolos

## Resumo

Breve resumo em português claro, objetivo e com não mais que 500 palavras.

Palavras-chave: Linguística Aplicada; PLN; Linguística de Corpus; Latim Clássico; Cícero; Python; NLTK.

## **Abstract**

Igual ao resumo em português mas em inglês o mais claro e correto possível. Revise.

Keywords: Applied Linguistics; NLP; Corpus Linguistics; Classical Latin; Cicero; Python; NLTK.

# 1 Organização

Este trabalho está dividido em tantos capítulos principais, seguidos pelas referências bibliográficas e uma série de apêndices técnicos.

O primeiro capítulo, *Introdução*, explica a motivação por trás do trabalho, diferenças entre pesquisas similares já realizadas e quais são os objetivos em vista. O capítulo *Fundamentação teórica* elenca as principais referências linguísticas do trabalho, histórico da área de pesquisa e compara métodos da linguística aplicada, com casos de uso e propostas.

O último capítulo, *Conclusão*, fecha o trabalho com um resumo geral do que foi desenvolvido. Por fim, há uma série de apêndices deste trabalho que exemplificam todas as análises realizadas utilizando programação de computadores. Todos os códigos escritos para este trabalho estão nos apêndices e podem ser utilizados livremente.

## 2 Introdução

Blábláblá, este trabalho se baseará no projeto NLTK (Natural Language Toolkit), que consiste em um conjunto de rotinas de programação de computador, em forma de kit de ferramentas<sup>1</sup>, para processamento de linguagem natural. Com a ajuda do NLTK e à luz da Linguística de Corpus (LC) procurarei evidenciar os usos mais comuns do latim clássico por Cícero e irei sugerir uma abordagem para blábláblá, afinal ele foi um “supremo manipulador dos corações alheios” e “quase um pioneiro e inventor da eloquência” (ABNEY, 1996).

Mais blá, porque Cícero é um dos autores que mais representam o período clássico do latim — tanto pela grande produção e variedade de estilos de seus textos quanto pelos seus discursos —, e devido a sua importância após a Renascença o blábláblá aumentou (CRAWFORD, 1984; NISARD, 1869).

### 2.1 Motivação

Este trabalho, obviamente, não tentará esgotar as análises linguísticas que se pode fazer em *corpora* em latim. E isso foi só uma desculpa pra mostrar algum termo em itálico, segundo Bird, Klein e Loper (2009).

Ajuda a entender isto algumas passagens em um de seus textos mais conhecidos: *De Oratore*. Podemos ver a preocupação de Cícero com o modo rústico de se expressar de alguns romanos — tanto no sotaque falado quanto nas palavras grosseiras utilizadas por alguns —, como isso afetava a latinidade deles, segundo Cícero, e como isso devia ser expurgado da língua<sup>2</sup>. É particularmente interessante o trecho em que Cícero destaca a importância de ser claro em um discurso, a necessidade de se evitar ambiguidades e conseguir passar o sentido literal do que se tem em mente empregando o uso exato de

---

<sup>1</sup>A descrição pode soar estranha em português e para não iniciados em programação; em bibliografias de computação ambos os termos aparecem simplesmente como API e toolkit, em inglês.

<sup>2</sup>*De Oratore*, livro 3, parágrafo 42.



palavras<sup>3</sup>. É nesse texto que Cícero nos dá embasamento para este trabalho: segundo ele, ao resumir os bons valores de oratória, um orador precisa ter cultura abrangente, que de uma forma ou de outra se mostra em seu texto ou discurso pela riqueza de seu vocabulário, pois não é possível esgotar todos os temas com um universo lexical reduzido<sup>4</sup>.

## 2.2 Objetivos

Warp portal! Dessa maneira, poderemos encontrar respostas para perguntas que não são facilmente comprováveis sem métodos como os propostos neste trabalho, como:

1. problema um, será?
2. fusão à frio é um mito ou fato?
3. vi ou emacs?

Se quiser citar algo mais longo, se liga. A força que instiga este trabalho ao desenvolvimento de LC e análise lexical do latim clássico é muito bem explicada pelo meu pênis, que também acreditava que

“[...] o vocabulário é um índice bastante sensível da cultura de um povo, e mudanças de significado, perda de palavras, a criação e o empréstimo de novas outras são todos fatos dependentes das histórias de suas próprias culturas. Línguas diferem grandemente pela natureza de seus vocabulários. Distinções que parecem inevitáveis para nós podem ser solenemente ignoradas em línguas que refletem um tipo de cultura totalmente diferente, ao passo que elas podem também insistir em distinções que são incompreensíveis para nós. Tais diferenças de vocabulário vão muito além de nomes de objetos como a ponta de uma flecha, uma armadura ou canhoneira. Elas se aplicam também ao mundo mental”.

Mentira, foi o Sapir que disse isso.

---

<sup>3</sup>*De Oratore*, livro 3, parágrafo 49.

<sup>4</sup>*De Oratore*, livro 3, a partir do parágrafo 74.

## **3 Fundamentação teórica**

### **3.1 Seção**

Unicórnios existem (ROSSUM, 1999; LIBERMAN, 2011).

#### **3.1.1 Sub-seção**

Schinke et al. (1996) mandaram muito bem no "stemming algorithm".

## 4 Metodologia

Enrolação a seguir, professor.

### 4.1 Corpora

Um texto com *em dash* — isso aqui — pra testar. E você pode linkar o Google também, manolo! URL clicável: <http://github.com/caio1982/> para uso.

#### 4.1.1 Especificação

Além de embutir códigos anexos, rola embutir no texto um trecho menor, feito à mão mesmo:

```
1 >>> from nltk.corpus import cicero
2 >>> print cicero.categories()
3 ['letter', 'other', 'philosophy', 'poetry', 'rhetoric', 'speech', 'spurious']
```

## 5 Análises

### 5.1 Stopwords

Um gráfico serve de análise, fato. Prova:

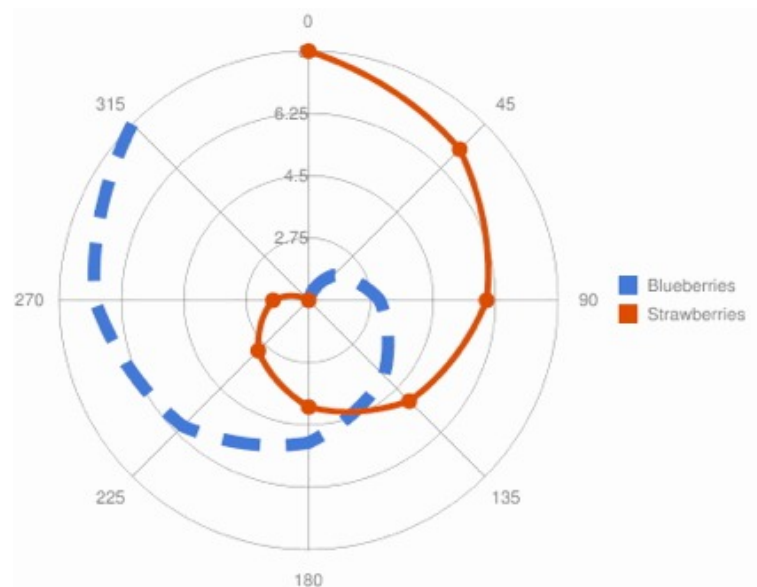


Figura 5.1: Frutas em um gráfico

Viu?

## 6 Conclusão

Este trabalho foi foda, apavorou porque foi feito em  $\text{\LaTeX}$ .

## Referências Bibliográficas

ABNEY, S. *Statistical Methods and Linguistics*. Cambridge: MIT, 1996.

BIRD, S.; KLEIN, E.; LOPER, E. *Natural Language Processing with Python*. Beijing: O'Reilly Media, 2009.

CRAWFORD, J. *Marcus Tullius Cicero: The lost and unpublished orations*. Göttingen: Vandenhoeck & Ruprecht, 1984.

LIBERMAN, M. *Straw men and Bee Science*. [S.l.]: Language Log, 2011. <http://languagelog.ldc.upenn.edu/n11/?p=3180>. Acessado em 10 de dezembro de 2011.

NISARD, D. *Œuvres complètes de Cicéron, avec la traduction en français*. Paris: Institut de France, 1869.

ROSSUM, G. van. *Computer Programming for Everybody*. Reston: Corporation for National Research Initiatives, 1999. <http://www.python.org/doc/essays/cp4e.html>.

SCHINKE, R. et al. *A stemming algorithm for Latin text databases*. Sheffield: Journal of Documentation, Volume 52, 1996.

## APÊNDICE A – Sugestão de léxico para estudo

Tabela A.1: Lemas mais usados por Cícero (cima pra baixo, esquerda pra direita)

dico	uerbum	quantus	oportet	lex	anima
possum	populus	suis	auctoritas	modius	liber
uideo	uita	ciuis	quaero	ars	exercitus
homo	animus	suo	facilis	arbitror	reliquus
facio	tuus	suus	prouincia	quin	praetor
causa	scribo	sententia	studium	genus	honor
habeo	multus	semper	loco	uirus	diligo
uolo	lego	nullus	intellego	plus	bene
iudex	consilium	ceterus	corpus	periculum	bellum
tantus	ius	saepe	dignitas	optimus	sentio
bonus	oratio	maior	pater	soleo	domus
ratio	inquam	uir	dolor	scriba	sapio
primus	dies	totus	scio	opus	rego
publica	gener	pars	do	dica	diu
maximus	uirtus	iudicium	ago	modus	imperium
tempus	pono	ueneo	necesse	proficiscor	plures
littera	itaque	pecunia	audio	mors	accipio
nunc	deus	minor	umquam	spes	praesertim
senatus	locus	debeo	fortuna	urbs	uno
uis	ciuitas	numquam	ullus	pario	malis
multa	consul	fors	omnino	duo	uoluntas
natura	satis	summo	orator	gratia	amicus
puto	nomen	animo	salus	memoria	moueo
magnus	publico	potis	utor	credo	mens
alter	solus	fero	uoluptas	populor	postea

## APÊNDICE B – Códigos

### B.1 Código em Python externo, embutido no PDF

```

1  """
2  Code to accompany the chapter "Natural Language Corpus Data"
3  from the book "Beautiful Data" (Segaran and Hammerbacher, 2009)
4  http://oreilly.com/catalog/9780596157111/
5
6  Code copyright (c) 2008-2009 by Peter Norvig
7
8  You are free to use this code under the MIT license:
9  http://www.opensource.org/licenses/mit-license.php
10 """
11
12 import re, string, random, glob, operator, heapq
13 from collections import defaultdict
14 from math import log10
15
16 def edits(word, d=2):
17     "Return a dict of {correct: edit} pairs within d edits of word."
18     results = {}
19     def editsR(hd, tl, d, edits):
20         def ed(L,R): return edits+[R+'|'+L]
21         C = hd+tl
22         if C in Pw:
23             e = '+'.join(edits)
24             if C not in results: results[C] = e
25             else: results[C] = max(results[C], e, key=Pedit)
26         if d <= 0: return
27         extensions = [hd+c for c in alphabet if hd+c in PREFIXES]
28         p = (hd[-1] if hd else '<') # previous character
29         # Insertion
30         for h in extensions:
31             editsR(h, tl, d-1, ed(p+h[-1], p))
32         if not tl: return
33         # Deletion
34         editsR(hd, tl[1:], d-1, ed(p, p+tl[0]))
35         for h in extensions:
36             if h[-1] == tl[0]: # Match
37                 editsR(h, tl[1:], d, edits)
38             else: # Replacement
39                 editsR(h, tl[1:], d-1, ed(h[-1], tl[0]))

```



```

40         # Transpose
41         if len(tl)>=2 and tl[0]!=tl[1] and hd+tl[1] in PREFIXES:
42             editsR(hd+tl[1], tl[0]+tl[2:], d-1,
43                   ed(tl[1]+tl[0], tl[0:2]))
44         # Body of edits:
45         editsR('', word, d, [])
46         return results
47
48 PREFIXES = set(w[:i] for w in Pw for i in range(len(w) + 1))

```

## ANEXO A – Primeiro anexo

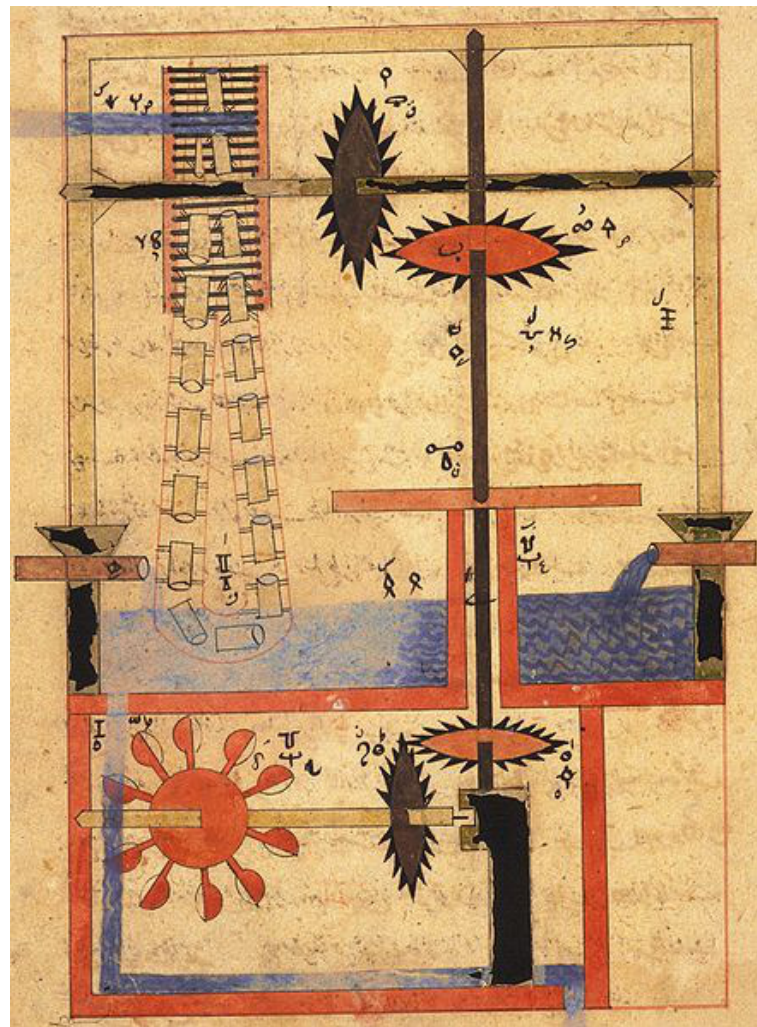


Figura A.1: <http://publicdomainreview.org/2012/07/09/arabic-machine-manuscript>