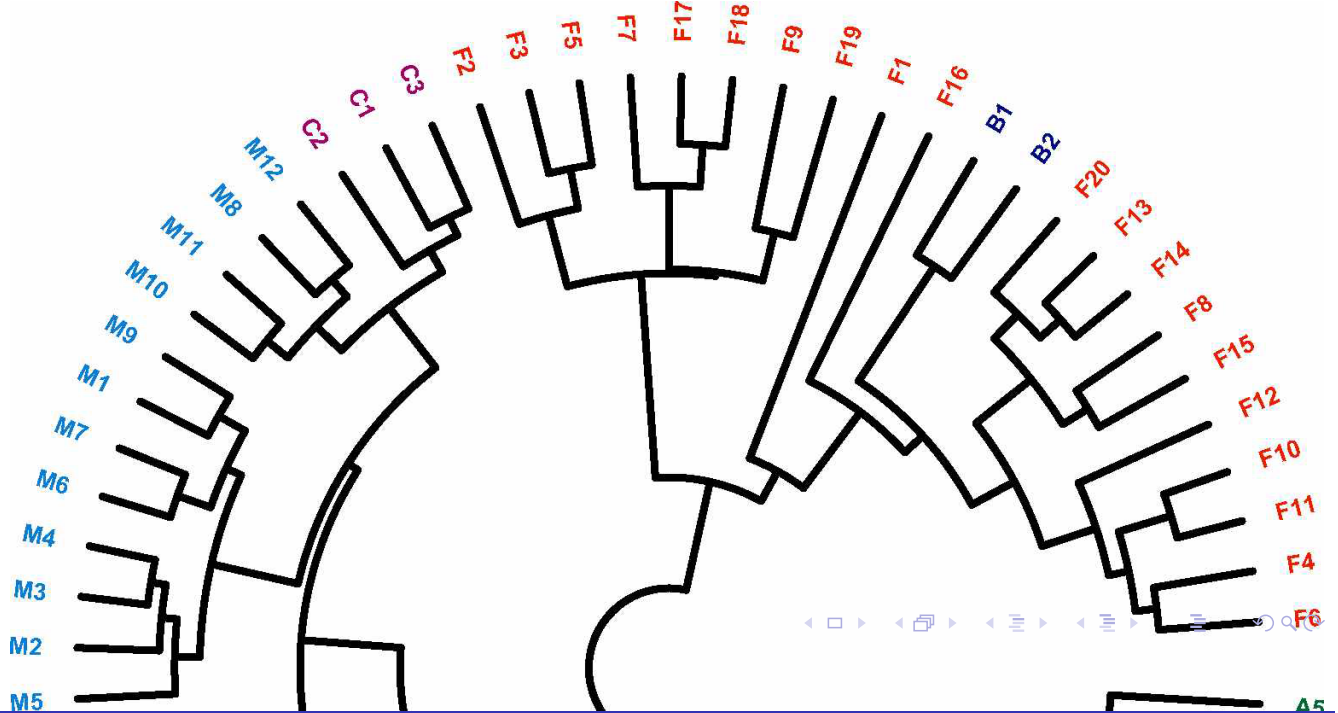


Um futuro risonho

Colaboração entre a Linateca e o NuPILL

Diana Santos

d.s.m.santos@ilos.uio.no



Colaboração com o NuPILL

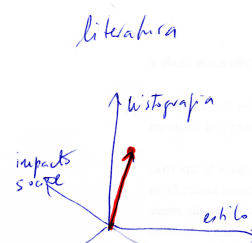
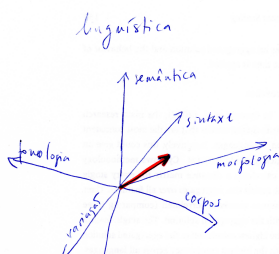
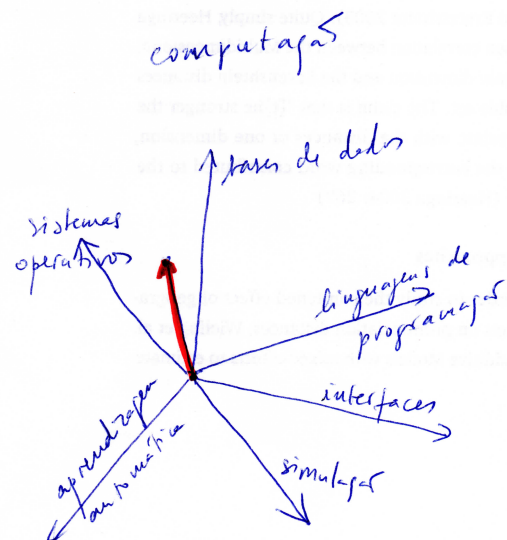
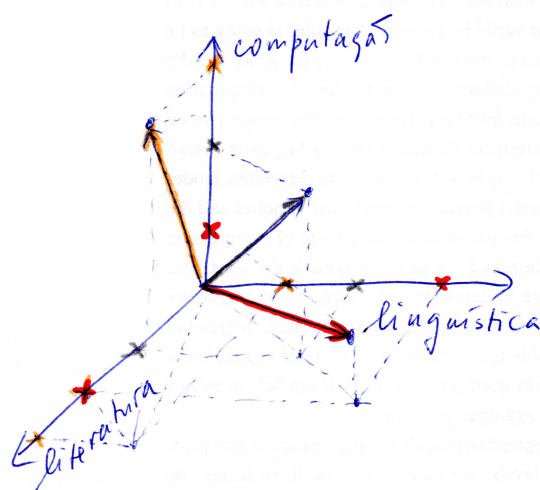
- Graças ao Emanuel Pires, que arranhou fundos da FAPEMA para ir a Oslo, e que convidou Alckmar dos Santos para uma palestra convidada na universidade, exatamente sobre o NuPILL
- Troca de ideias para futuras colaborações no *Primeiro Encontro sobre leitura distante em português*, experiência de integração de um texto do NuPILL na Linateca, e da análise da Linateca para um texto do NUPILL
- Tentativa de colaboração institucional ao nível do ensino (da literatura em português), em curso
- Esta palestra/conversa, para tentarmos assentar ideias

O NuPILL e a Linguateca: projetos iguais?

O **NUPILL** tem como proposta central o desenvolvimento de pesquisas sobre os textos literários em meio digital, abrangendo uma gama que vai da criação à leitura utilizando dispositivos e processos da informática, passando por estratégias e ferramentas de ensino e aprendizagem de literatura.

O objectivo da **Linguateca**, um centro de recursos – distribuído – para o processamento computacional da língua portuguesa, é servir a comunidade que se dedica ao processamento da nossa língua. Em particular, a) facilitar o acesso aos recursos já existentes, através do desenvolvimento de serviços de acesso na rede, e mantendo um portal com informação útil; b) desenvolver, de forma harmoniosa, em colaboração com os interessados, os recursos considerados mais prementes; c) organizar avaliações conjuntas que envolvam a comunidade como um todo.

Interdisciplinaridade



... o sonho é uma constante da vida ...

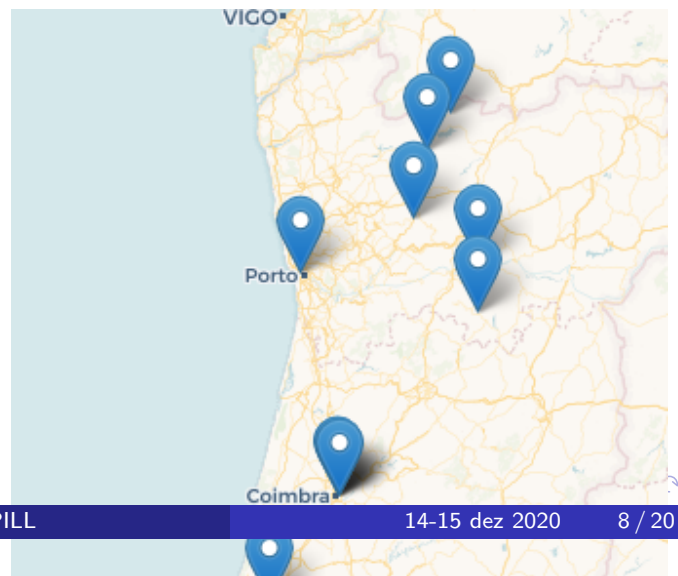
A Pedra Filosofal, António Gedeão



- Em termos de recursos, o dito no slide anterior
- Mas em termos de pesquisa, abrem-se novos horizontes conjuntos
 - “colaboração/diálogo” entre ferramentas
 - desenvolvimento de novas ferramentas conjuntas
 - estudos comuns ou paralelos
 - criação de avaliações conjuntas sobre literatura

Algumas veredas

- Visão cruzada: as descrições do Brasil na literatura portuguesa, e as descrições de Portugal na literatura brasileira
- Intertextualidade e influência(s) estrangeira(s)
- A casa brasileira e a casa portuguesa: a sua localização nos romances
- A flora e a fauna; as profissões; a religião
- Personagens e nacionalidades
- Festas e costumes



A ação “Distant reading for European literary history”

Distant Reading

This action will affect the way scholars in the Humanities do research, but also the way institutions like libraries will make their holdings available to researchers in the future

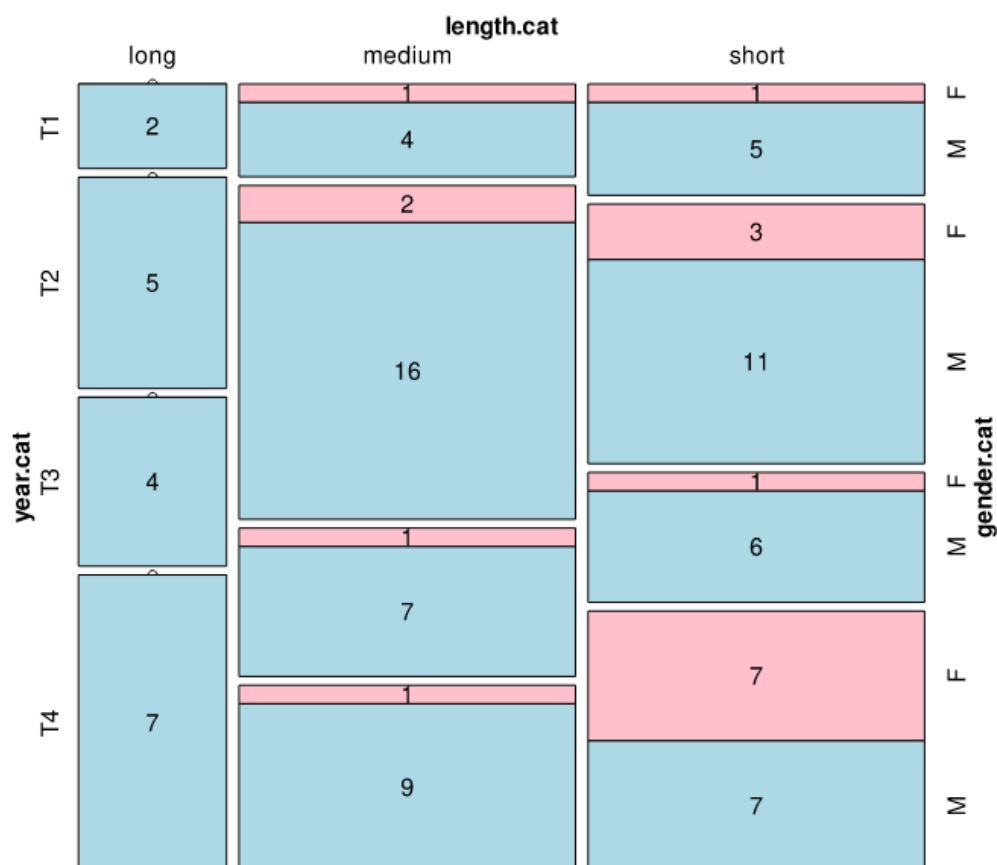
- http://www.cost.eu/COST_Actions/ca/CA16204
- <https://www.distant-reading.net/>
- <https://distantreading.github.io/>

A coleção ELTec-por e as outras coleções:

<https://distantreading.github.io/ELTeC/>: 1840-1919, romances portugueses. ELTeC-por-ext: também inclui romances brasileiros do mesmo período.

Navigation icons: back, forward, search, etc.

A coleção ELTeC-por através de alguns parâmetros



Navigation icons: back, forward, search, etc.

Natal na coleção ELTeC-por

- E da Páscoa? Ainda menos, 19 ocorrências em 15 obras. E do Carnaval? 38 em 10 obras (2 com cenas nesse período).

Títulos de obras na coleção ELTeC-por

A black and white photograph of a woman with long, dark, wavy hair, smiling at the camera. She is wearing a light-colored, sleeveless dress with a dark, patterned bodice. The background features a large potted plant with broad leaves on the left and a wall with a framed picture on the right.

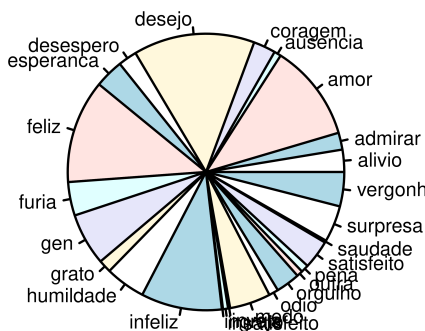
Emoções por autores

Ordenados pela data de nascimento, e indicando número de obras na Literateca, número de palavras, número de emoções, e número relativo de emoções por número de palavras.

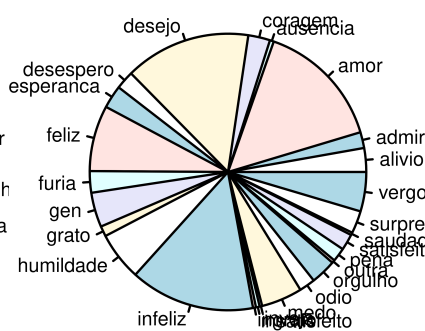
CamCBra	JulDin	MacAss	EcaQue	CoeNet	RauBra
1825-1890	1839-1871	1839-1908	1845-1900	1864-1934	1867-1930
60	19	162	28	15	16
3.446.039	1.086.096	2.208.930	2.773.552	763.541	722.768
111.233	34.001	56.150	69.800	21.510	18.360
0,032	0,031	0,025	0,025	0,028	0,025

Emoções por autores

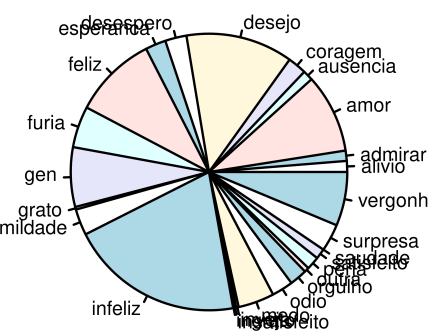
Júlio Dinis



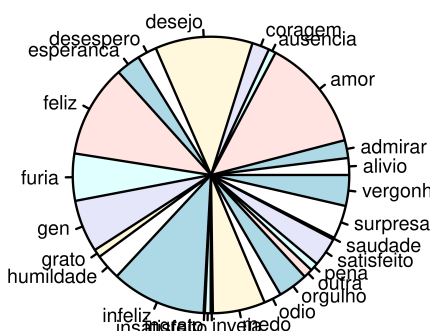
Camilo Castelo Branco



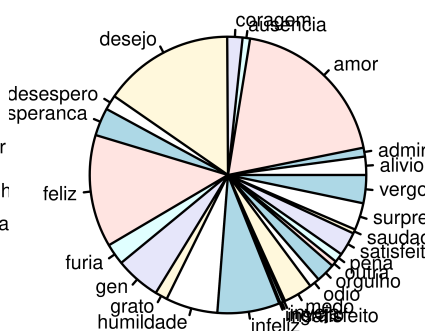
Raul Brandão



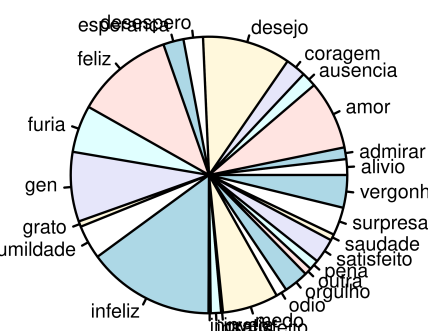
Eça de Queirós



Machado de Assis



Coelho Neto



Avaliação conjunta

Um processo para tentar convencer a comunidade a juntar-se e a definir uma tarefa relevante, neste caso para a leitura distante da literatura em português.

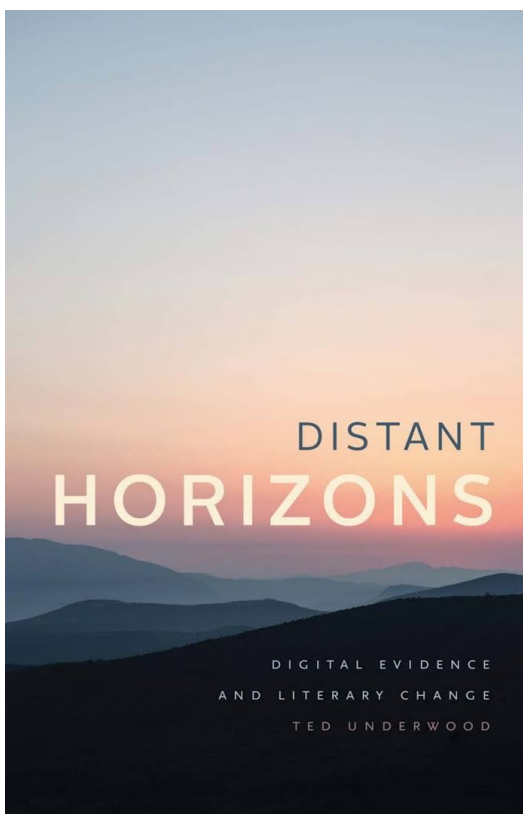
- Sistemas que identificam as personagens principais de romances
- Sistemas que identificam a localização das cenas dos romances
- Sistemas que detectam o tipo de narrador
- Sistemas que medem o tempo descrito no romance

Outras ideias?



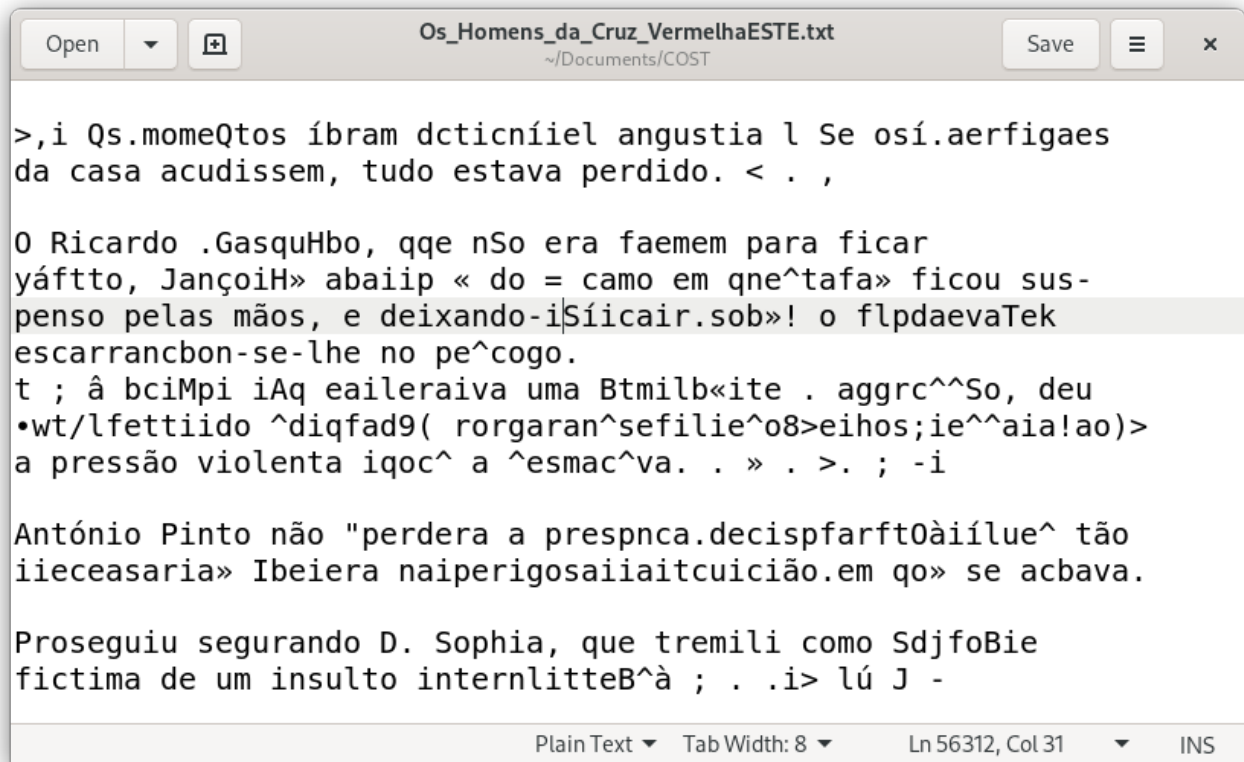
Corrida de canoa -ubatuba

Melhorar o ROC (reconhecimento ótico de caracteres)



Algo que pode convencer a Google?

- Temos um conjunto enorme de obras mal reconhecidas
- Temos um conjunto de obras já revistas
- Que tal fazer um programa de ROC muito melhor?
- Que tal tentar também fazer (separadamente) um modernizador?



Os_Homens_da_Cruz_VermelhaESTE.txt
~/Documents/COST

>,i Qs.momeQtos íbram dcticníiel angustia l Se osí.aerfigaes
da casa acudissem, tudo estava perdido. < . ,

O Ricardo .GasquHbo, qqe nSo era faemem para ficar
yáftto, JançoiH» abaiip « do = camo em qne^tafa» ficou sus-
penso pelas mãos, e deixando-i\$íicair.sob»! o flpdaevaTek
escarrancbon-se-lhe no pe^cogo.
t ; â bciMpi iAq eaileraiva uma Btmilb«ite . aggrc^^So, deu
•wt/lfettiido ^diqfad9(rorgaran^sefilie^o8>eihos;ie^^aia!ao)>
a pressão violenta iqoc^ a ^esmac^va. . » . >. ; -i

António Pinto não "perdera a prespnca.decispfarft0àiiíluue^ tão
iieceasaria» Ibeiera naiperigosaiiaitcuicição.em qo» se acabava.

Proseguiu segurando D. Sophia, que tremili como SdjfoBie
fictima de um insulto internlitteB^à ; . .i> lú J -

Plain Text ▾ Tab Width: 8 ▾ Ln 56312, Col 31 ▾ INS

Um futuro risonho?

- Finais felizes?
- Um futuro risonho? (6 casos)
- Ou ri melhor quem ri por último? (1 caso)

Que estudos de literatura
podemos criar em
conjunto?

Procura: [lema="futuro"] @[pos="ADJ.*"]

Distribuição de **lema**

Corpo: Literateca v. 5.16

217 casos.

Distribuição

Houve **132** valores diferentes de **lema**.

bom	12
brilhante	12
incerto	10
próspero	7
glorioso	7
risonho	6
próximo	5
certo	5
remoto	4
perdido	4
ridente	3
tranquilo	3
distante	3

solitario 2

- <https://manifesto74.pt/as-venias-da-republica/>
- <https://www.dn.pt/edicao-do-dia/06-jul-2019/o-nosso-rainha-da-inglaterra-11067977.html>
- <https://novaimpressa.com/2017/09/corrida-de-canoa-caicara-completa-60.html>

Obrigada!

