

Anexo B: Seleção de conteúdo por ranqueamento (e.g. por tamanho de palavras) e limiar (e.g. número mínimo de palavras)

importante bibliotecas úteis:

```
In [54]: from SPARQLWrapper import SPARQLWrapper, JSON
import time, numpy as n, nltk as k
```

Definição de prefixos úteis para as triplas rdf:

```
In [55]: PREFIX=""PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ops: <http://purl.org/socialparticipation/ops#>
PREFIX opa: <http://purl.org/socialparticipation/opa#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX tsioc: <http://rdfs.org/sioc/types#>
PREFIX schema: <http://schema.org/>""
```

Buscando todos os comentários no endpoint SparQL disparado pelo Jena:

```
In [56]: NOW=time.time()
q="SELECT ?comentario ?titulo ?texto WHERE \
      {?comentario dc:type tsioc:Comment.\
      OPTIONAL {?comentario dc:title ?titulo . }\
      OPTIONAL {?comentario schema:text ?texto .}}"
sparql3 = SPARQLWrapper("http://localhost:82/participabr/query")
sparql3.setQuery(PREFIX+q)
sparql3.setReturnFormat(JSON)
results4 = sparql3.query().convert()
print("%.2f segundos para puxar todos os comentários do Participa.br"%(time.time()-NOW,))
```

2.34 segundos para puxar todos os comentários do Participa.br

Limpando mensagens das sugeiras do BD:

```
In [57]: msgs=results4["results"]["bindings"]
msgs=[mm for mm in msgs_ if ("titulo" not in mm.keys()) or (("teste de stress" not in mm["titulo"]["value"].lower())
and ("comunidade de desenvolvedores e nesse caso, quanto mais" not in mm["texto"]["value"].lower()))]
```

Fazendo função para extrair atributos das mensagens:

```
In [58]: import string, numpy as n
exclude = set(string.punctuation+u'\u201c'+u'\u2018'+u'\u201d'+u'\u2022'+u'\u2013')
def atributos(_msg):
    texto=_msg["texto"]["value"]
    texto_ = ''.join(ch for ch in texto if ch not in exclude)
    palavras=texto_.split()
    tams=[]
    for palavra in palavras:
        tams.append(len(palavra))
    return len(tams), n.mean(tams), n.std(tams)
```

Criando vetor de atributos de cada mensagem:

```
In [59]: atrs=[]
for msg in msgs:
    atrs.append(atributos(msg))
atrs=n.array(atrs)
```

Fazendo seleção das mensagens que possuem entre 110 e 115 mensagens:

```
In [60]: max_palavras=115
min_palavras=110
n_msgs=((atrs[:,0]>min_palavras)*(atrs[:,0]<max_palavras)).sum()
print(u"são %i mensagens com mais de %i palavras e menos de %i"%
      (n_msgs, min_palavras, max_palavras) )
```

são 26 mensagens com mais de 110 palavras e menos de 115

Selecionando mensagens com média do tamanho da palavra maior que 6 caracteres:

```
In [61]: msgs_i=((atrs[:,0]>min_palavras)*(atrs[:,0]<max_palavras)).nonzero()[0]
textos=[msgs[i]["texto"]["value"] for i in msgs_i[j]
        for j in (atrs[msgs_i][:,1]>6.).nonzero()[0]]
```

```
for texto in textos: print texto+"\n=====
```

1. Auxiliar na divulgação das informações jurídicas para estrangeiros por meio de projetos de longo prazo dentro no âmbito das próprias universidades (ex. immigration/refugee law clinics), estimulando-se a parceria com as instituições, organizações e programas/projetos relacionados à temática migratória;
2. Inserir na nova política migratória a capacitação/formação específica para agentes públicos (obrigatória) que atuam ou possam vir a atuar em razão de suas atribuições institucionais com migrantes, refugiados e apátridas. Nesse sentido, a experiência e contribuição da academia será fundamental do ponto de vista da expertise científica (ensino e pesquisa - colaboração para o conteúdo programático da capacitação), além de reforçar a importância do ensino do Direito Internacional e Direitos Humanos nas Universidades.

=====

ROSAS, Agostinho da Silva; MELO NETO, José Francisco de. Educação Popular: Enunciados Teóricos. João Pessoa: Editora Universitária da UFPB, 2008.

Os autores buscam delimitar um campo da educação em que seja possível delinearem-se características que apenas à educação popular façam parte. Trata-se de um esforço de apresentação de constituintes com dimensões formadoras, tendo como aspectos centrais as experiências de vários educadores populares e as reflexões em desenvolvimento na época. Nessa perspectiva, são abordadas a experiência histórica, a cultura, o popular, a realidade, o trabalho, a autonomia, a liberdade e a igualdade como componentes fundantes para a realização de práticas em educação popular, lastreados pela dimensão ética do diálogo.

Disponível em:

http://www.prac.ufpb.br/copac/extelar/producao_academica/livros/pa_l_2004_educacao_popular_enunciados_teoricos.pdf

=====

Selecionando mensagens com média do tamanho da palavra menos que 5 caracteres:

```
In [62]: textos=[msgs[i]["texto"]["value"] for i in msgs_i][j]
         for j in (attrs[msgs_i][:,1]<5).nonzero()[0]
         for texto in textos: print texto+"\n=====
```

Seria suave aos ouvidos ver que a população participando e contribuindo pode notar que seus desejos desde que na medida do possível estão sendo ouvidos e colocado em prática.

A população cansou, pois até coloca sua cara pra bater, porém na hora de colocar em prática o que tanto desejam nada acontece.

O que resta, a saber, se quando a população participar da construção das políticas públicas, deixara suas vontades o campo da teoria para o campo da prática, muito difícil isso ocorrer ou talvez impossível, pois o desejo da população praticamente em todos as vezes e momentos que é aferida suas vontades, são deixados de lado, isso é uma pena.

=====

O §1º é a atividade ou ação natural, o principio de ofício ou por assim dizer a o fato singular do GT.

O §2º é a preposições facultadas que podem ou não advir, em conformidade das necessidades observadas na decorrência do §1º, ou seja, dependente do §1º.

O §3º é mais fácil, pois não há registros do que não aconteceu, então a referencia, por lógica, deve ser a ultima.

Não me ative em pensar em mesmo método com os incisos, por serem muitos e não existe espaço para estabelecer a desfiguração, observando que o contexto das determinações estarem em elevação de importância, mesmo que a ordenação ou 'cronologia' traga maior facilidade de compreensão.

=====

Laura, essa alteração legislativa é muito importante. É muito desproporcional exigir que o migrante tenha que indicar uma repartição consular em outro país para obter o visto. O fundamento está no fato de que o Estatuto do Estrangeiro não permite concessão de visto a quem está irregular. Mas é possível uma interpretação sistemática e razoável da norma, para que seja feita uma distinção para os estrangeiros indocumentados (que estão ainda sem o visto mas que têm o direito à regularização migratória, como por exemplo por motivo de reunião familiar). Exigir que o estrangeiro que já está no Brasil tenha que sair do país apenas para buscar o visto é muito oneroso e desproporcional.

=====

Fazendo contagem das palavras mais frequentes para seleção:

```
In [63]: NOW=time.time()
import string, nltk as k
palavras=string.join([i["texto"]["value"].lower() for i in msgs])
exclude = set(string.punctuation+u'\u201c'+u'\u2018'+u'\u201d'+u'\u2022'+u'\u2013')
palavras = ''.join(ch for ch in palavras if ch not in exclude)
palavras_=palavras.split()
print(u"feita lista de todas as palavras de todos os comentários em %.2f"%(time.time()-NOW,))
stopwords = set(k.corpus.stopwords.words('portuguese'))
palavras_=[pp for pp in palavras_ if pp not in stopwords]
fdist=k.FreqDist(palavras_)
print("retiradas stopwords feita contagem das palavras em %.2f"%(time.time()-NOW,))
```

feita lista de todas as palavras de todos os comentários em 0.17

retiradas stopwords feita contagem das palavras em 0.45

Fazendo seleção das 14 palavras mais incidentes nos comentários do Participa.br:

```
In [65]: for fd,ii in [(fdist[i],i) for i in fdist.keys()[:14]]: print fd, ii
```

1277 é
1256 não
762 ser
717 participação
548 social
526 sociedade
468 à

7/28/2014

2. Seleção por ranqueamento e limiar

459 sobre
367 governo
357 são
337 forma
327 políticas
310 públicas
302 brasil

|||-- FIM --||