

Reunião 09/03/20

Objetivos propostos no dia 17/02

- Implementação e realização de testes na arquitetura criada pelo co-orientador Ricardo Ribeiro através da implementação de novos modelos NLP estudados anteriormente .
 - Este objetivo foi alterado para a reconstrução de uma arquitetura totalmente nova no dia 21/02.

RESULTADOS OBTIDOS DA ARQUITETURA CRIADA

(utilizando o modelo mais pequeno e rápido do SpaCy)

```
<topic>
  <id>001</id>
  <uid>u1</uid>
  <title>Having Beers in a Bar</title>
  <description>Find the moment in 2015 and 2016 when u1 enjoyed
beers in the bar.</description>
  <narrative>To be considered relevant, u1 must be clearly in a
bar. Any moments that u1 drinks beers at home or outside without
the bar view are not considered relevant.</narrative>
</topic>
```

```
Relevant things ['beers']
Activities ['having beers', 'enjoyed beers']
Dates ['2015', '2016']
Locations: ['bar']
Location or Thing or Date []
User inside: True
User outside: False

Other words: ['having', 'u1', 'enjoyed', 'clearly']
other things []
```

```
<topic>
  <id>002</id>
  <uid>u1</uid>
  <title>Building Personal Computer</title>
  <description>Find the moment when u1 built my personal
computer from scratch.</description>
  <narrative>To be considered relevant, u1 must be clearly at
the office with the PC parts on the table. Any moments that the u1
is not in the office or there are no PC parts/uncompleted PCs on
the table are not considered relevant.</narrative>
</topic>
```

```
Relevant things ['personal computer', 'computer', 'scratch', 'pc parts', 'parts', 'table']
Activities ['building personal computer']
Dates []
Locations: ['office']
Location or Thing or Date []
User inside: True
User outside: False

Other words: ['building', 'personal', 'u1', 'built', 'clearly']
other things ['pc']
```

```
<topic>
  <id>003</id>
  <uid>u1</uid>
  <title>In A Toy Shop</title>
  <description>Find the moment when u1 was looking at items in
a toyshop.</description>
  <narrative>To be considered relevant, u1 must be clearly in a
toyshop. Various toys are being examined, such as electronic
trains, model kits and board games. Being in an electronics store,
or a supermarket, are not considered to be relevant.</narrative>
</topic>
```

```
Relevant things ['items', 'various toys', 'toys', 'electronic trains', 'trains', 'model kits', 'board games']
Activities ['looking at items', 'toys are being examined']
Dates []
Locations: ['toy shop', 'toyshop']
Location or Thing or Date []
User inside: True
User outside: False

Other words: ['u1', 'looking', 'clearly', 'examined', 'electronic']
other things ['toy', 'shop', 'model', 'kits', 'board', 'games']
```

```
<topic>
  <id>004</id>
  <uid>u1</uid>
  <title>Television Recording</title>
  <description>Find the moments when u1 was being recorded for
a television show.</description>
  <narrative>To be considered relevant, there must clearly be a
television camera in front of u1. The moments the
interviewer/cameramen is interviewing/recording u1 are also
considered relevant. This can take place at home or in another
location. All recording took place in one day and in more than one
location.</narrative>
</topic>
```

```
Relevant things ['television show', 'television camera', 'interviewer', 'cameramen', 'place']
Activities ['television recording', 'recorded for a television', 'being recorded for a television show', 'interviewing', 'recording']
Dates []
Locations: ['home', 'location']
Location or Thing or Date []
User inside: True
User outside: False

Other words: ['recorded', 'clearly', '/', 'considered', 'relevant', 'took']
other things ['television', 'camera', 'moments', 'day']
```

```
<topic>
  <id>005</id>
  <uid>u1</uid>
  <title>Public Transport In Home Country</title>
  <description>Find the moments in 2015 and 2018 when u1 was
using      public      transports      in      my      home      country
(Ireland).</description>
  <narrative>Taking any form of public transport in Ireland is
considered relevant, such as bus, taxi, train and boat. The
moments that u1 is driving a car is not relevant</narrative>
</topic>
```

```
Relevant things ['public transport', 'transport', 'public transports', 'transports', 'home', 'form', 'bus', 'taxi', 'train', 'boat']
Activities ['using public transports']
Dates ['2015', '2018']
Locations:  ['home country', 'ireland']
Location or Thing or Date []
User inside: Not known
User outside: Not known

Other words:  ['public', 'u1', 'taking', 'considered', 'relevant']
other things ['country']
```



```
<topic>
  <id>006</id>
  <uid>u1</uid>
  <title>Seaside Moments</title>
  <description>Find moment(s) in which u1 was walking by the
sea taking photos or eating ice cream.</description>
  <narrative>To be considered relevant, u1 must be taking a
walk by the sea or eating ice cream and the sea is clearly
visible. u1 must also be shown taking a photo with his/her camera
phone or holding the ice cream cone at this time. On the beach or
walking on a pier are both considered relevant once u1 is taking a
photo or eating ice cream.</narrative>
</topic>
```

```
Relevant things ['photos', 'ice cream', 'cream', 'walk', 'photo', 'camera', 'camera phone', 'ice cream cone', 'cream cone', 'cone', 'beach', 'pier']
Activities ['walking by the sea', 'taking photos', 'eating ice cream', 'taking a walk', 'taking a photo', 'holding the ice cream cone', 'walking on a pier']
Dates []
Locations: ['sea']
Location or Thing or Date []
User inside: False
User outside: True

Other words: ['u1', 'walking', 'taking', 'eating', 'clearly', 'visible', 'shown', '/', 'holding', 'considered', 'relevant']
other things ['seaside', 'moments', 'ice', 'phone', 'time']
```

```
<topic>
  <id>007</id>
  <uid>u1</uid>
  <title>Grocery Stores</title>
  <description>Find moment(s) in 2016 and 2018 when u1 was
grocery shopping on the weekends.</description>
  <narrative>To be considered relevant, u1 must be clearly in a
grocery shop and buy something from it. Any moments on Saturdays
and Sundays when u1 was in a grocery store and visibly interacting
with products are also considered relevant.</narrative>
</topic>
```

```
Relevant things ['grocery stores', 'something', 'products']
Activities ['grocery shopping', 'interacting with products']
Dates ['2016', '2018', 'weekends', 'saturdays', 'sundays']
Locations: ['grocery shop', 'grocery store']
Location or Thing or Date []
User inside: True
User outside: False
```

```
Other words: ['find', 'moment(s)', 'u1', 'clearly', 'buy', 'visibly', 'interacting', 'considered', 'relevant']
other things ['grocery', 'stores', 'shopping', 'shop', 'moments', 'store']
```

```
<topic>
  <id>008</id>
  <uid>u1</uid>
  <title>Photograph of The Bridge</title>
  <description>Find the moment when u1 was taking a photo of a
bridge.</description>
  <narrative>Moments when u1 was walking on a street without
stopping to take a photo of a bridge are not relevant. Any other
moment showing a bridge when a photo was not being taken are also
not considered to be relevant.</narrative>
</topic>
```

```
Relevant things ['photograph', 'photo']
Activities ['taking a photo']
Dates []
Locations: ['bridge']
Location or Thing or Date []
User inside: True
User outside: False

Other words: ['u1', 'taking']
other things []
```

```
<topic>
  <id>009</id>
  <uid>u1</uid>
  <title>Car Repair</title>
  <description>Find the moment when u1 was repairing his car in
the garden.</description>
  <narrative>Moments when u1 was repairing his car in the
garden with the gloves on his hand. Sometimes, he also held the
hammer and his phone and these moments are also considered
relevant. The moments when u1 was with the cars without showing
repair action are not considered relevant.</narrative>
</topic>
```

```
Relevant things ['car repair', 'car', 'gloves', 'hand', 'hammer', 'phone']
Activities ['repairing his car']
Dates []
Locations: ['garden']
Location or Thing or Date []
User inside: True
User outside: False

Other words: ['u1', 'repairing', 'held', 'considered', 'relevant']
other things ['repair', 'moments']
```

```
<topic>
  <id>010</id>
  <uid>ul</uid>
  <title>Monsters</title>
  <description>Find the moment(s) when ul was looking at an old
clock, with flowers visible, with a small monster watching
ul.</description>
  <narrative>Moments when ul was at home, looking at an old
clock, with flowers visible, with a lamp and perhaps two small
monsters watching ul are considered relevant. One of the monsters
might be a long rabbit. The moments without one of the
aforementioned conditions: monsters, flowers, and old clock are
not considered relevant.</narrative>
</topic>
```

```
Relevant things ['old clock', 'clock', 'flowers', 'flowers visible', 'small monster', 'monster', 'lamp', 'small monsters', 'monsters', 'long rabbit', 'rabbit']
Activities ['looking at an old clock']
Dates []
Locations: ['home']
Location or Thing or Date []
User inside: True
User outside: False

Other words: ['s', 'looking', 'old', 'visible', 'small', 'watching', 'considered', 'relevant', 'long']
other things ['moments']
```

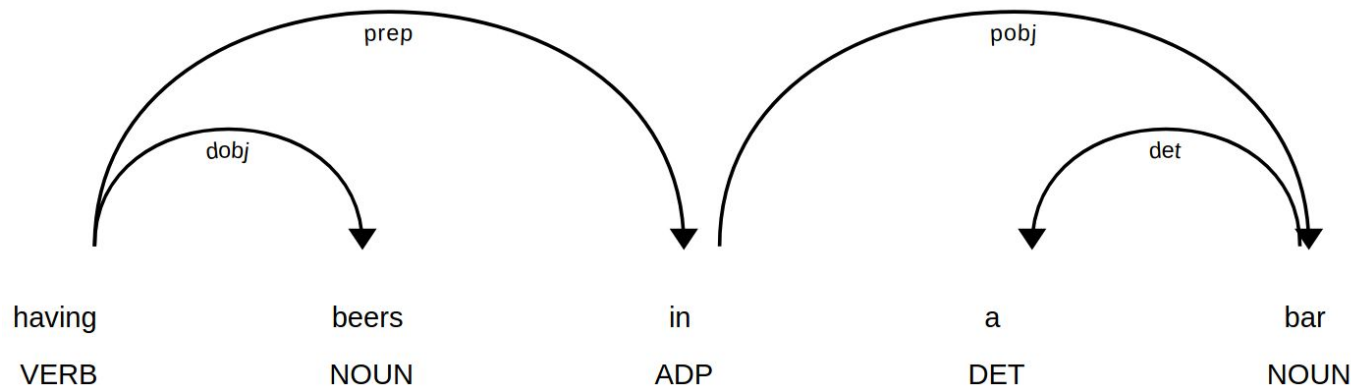
Algumas ideias implementadas no pré-processamento

- O tópico da narrativa foi repartido em várias frases, onde o ponto de split corresponde aos pontos finais.
- As frases foram inicialmente convertidas para letras minúsculas.
- Frases que acabavam em “not considered to be relevant” , “is not relevant” e semelhantes, foram descartadas e adicionadas a “Frases negativas”, para tentar implementar uma nova categoria **(explicado mais à frente)**
- Frases iniciadas com “Find the moments”, “Find the moment”, etc, foram pré-processadas de forma a remover essas mesmas palavras.

Algumas ideias básicas implementadas no algoritmo

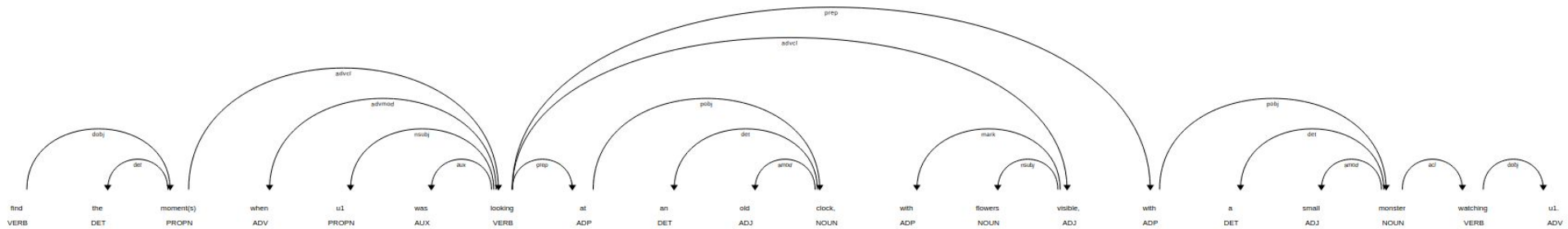
- Se a palavra for um dígito é considerada uma data.
- Se a PoS (Part of Speech) da palavra for um “ADJ” a palavra que vem a seguir é algo relevante.
- Se a dependência da palavra for um “compound” é porque a palavra a seguir é algo relevante (ex: camera phone) ou uma atividade caso tenha um verbo anteriormente (ex: eating ice cream).
- Se a palavra acabar em “ing” é porque corresponde a uma atividade
- Se a PoS for um “PRON” e a DeP um “poss” é porque a palavra a seguir é algo relevante

Alguns exemplos de como o algoritmo funciona com
a ajuda do Visualizer do SpaCy

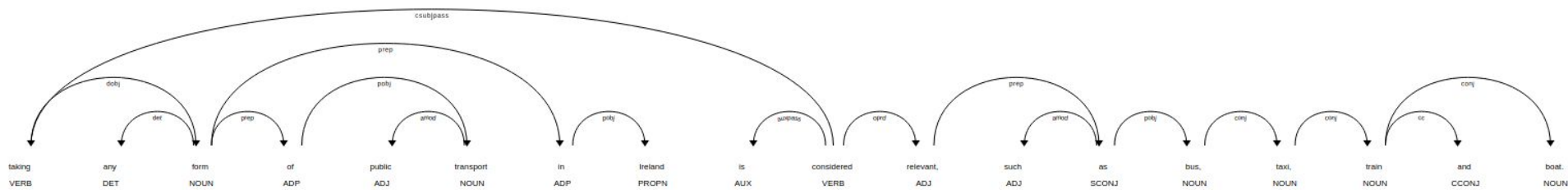


-> Para detetar o "Bar", o programa primeiramente deteta a palavra que corresponde a um "ADP", neste caso "in". Seguidamente, se a seguir ao ADP vier um DET e uma dependência "pobj" podemos afirmar que a palavra na posição de position + 2 corresponde a uma localização. Sendo que posição = posição da palavra "in"

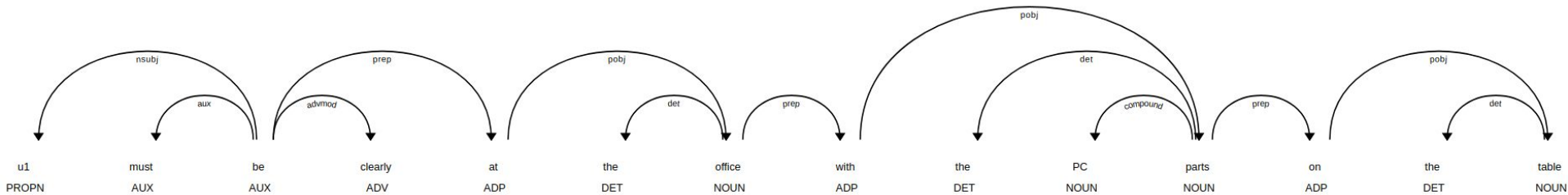
-> Para detetar a atividade o programa deteta a palavra "having" porque é um VERB e acaba em "ing". A seguir, junta a palavra a seguir caso esta seja um nome e um "dobj", neste caso "beers" o que leva a também a adicionar "beers" a algo relevante.



- “Looking at an old clock” dá-nos várias informações:
 - A palavra “Looking” diz-nos que vem uma atividade.
 - O facto de old ser um adjetivo é porque provavelmente se refere a uma coisa e não a uma localidade.
 - O facto do “ADP” ser “at” e de seguida vir um “an” dá-nos mais certezas que “old clock” é algo relevante e não uma localidade.



- Para retirar as palavras “bus”, “taxi”, “train” e “boat” é utilizada uma ferramenta do SpaCy denominada de “chunks”.
 - O facto de existir vírgulas entre “bus”, “taxi”, “train” e “boat” permite que estas palavras se tornem chunks de apenas uma palavra, o que facilita o processo de as procurar, e caso a dependência seja “pobj” ou “conj” adiciona a coisas relevantes.
- “Ireland” é retirada como localização porque é um “PROPN” que tem como dependência “pobj” e anteriormente tem um “ADP” que corresponde a “in”.
- “Public transport” é adicionado a coisas relevantes porque corresponde a um “ADJ” ligado a um “NOUN” que tem “pobj” como dependência



- “Office” é retirado como localização porque é um “NOUN” com dependência “pobj” em que anteriormente a ele tem um “ADP” e um “DET”, em que o “ADP” “corresponde a “at”, o que normalmente referências localizações ou objetos.
- “Table” é retirado como coisa relevante porque é um “NOUN” com dependência “pobj” em que anteriormente a ele tem um “ADP” e um “DET”, em que o “ADP” corresponde a “on”, o que só é utilizado para referenciar coisas.

Ideias por implementar

- **Palavras negativas** : Palavras que caso estejam presentes na imagem são automaticamente descartadas. Para isto as frases que foram descartadas inicialmente por acabarem em palavras semelhantes como “is not relevant” seriam processadas para esta categoria.
 - “The moments that u1 is driving a car is not relevant”. Teria interesse marcar “car” como palavra negativa e “driving a car” como atividade negativa.
 - “Being in an eletronics store, or a supermarket, are not considered to be relevant” - Teria intresse marcar “eletronics store” e “supermarket” como localização negativa.
- **Problema com esta ideia:**
 - “The moments without one of the aforementioned conditions : monsters, flowers, an old clock are not considered relevant”
 - Neste caso “monsters”, “flowers” , “old clock” **não são palavras negativas**

Solução idealizada a 04/03/20

- Comparar palavras negativas com as palavras que já foram detetadas anteriormente. Caso já tenham sido detectadas , continuam como palavras relevantes, caso não tenham sido é porque são palavras negativas.
- Isto resolveria o problema porque: “old clock”, “flowers” e “monsters” já tinham sido detetadas nas frases que não foram descartadas enquanto que “eletronic store”, “supermarket”, “cars” ainda não tinham sido.

FIM