

# LALIC

Laboratório de Linguística e Inteligência Computacional

## Text-image alignment in Portuguese news using LinkPICS



UNIVERSIDADE FEDERAL  
DE SÃO CARLOS

1

# Introduction

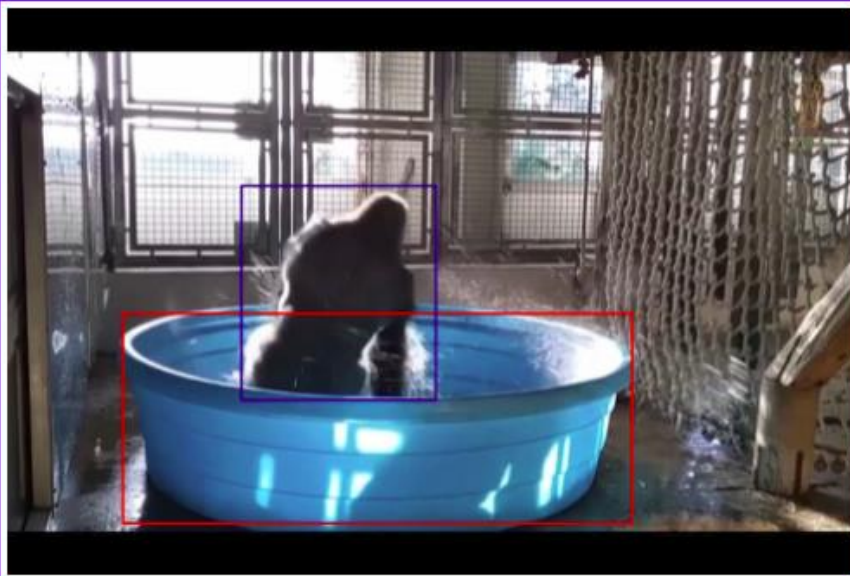


## What is Text-image alignment?

- Find correspondences between text and visual elements
- Can help on text understanding



# What is Text-image alignment?



## Título

Gorila dançarino impressiona com movimentos em **piscina**

## Legenda

Ele é uma das principais atrações do zoológico de Dallas  
Foto. BBC

## Texto

Em vídeo divulgado pelo zoo, o **animal** de 14 anos aparece recentemente dentro de uma **piscina** azul fazendo movimentos frenéticos comparados a passos de breakdance, um estilo de dança de rua. Zola já tinha viralizado antes, em 2011, quando imagens dele

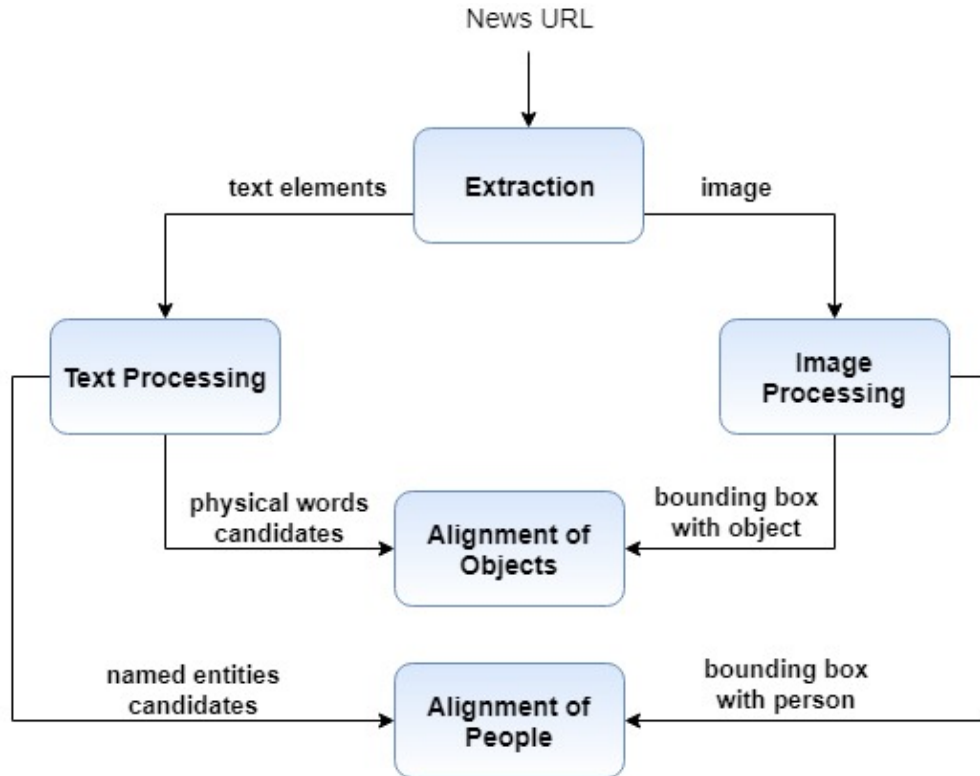
Figure 1: Text-image alignment performed by LinkPICS aligner

# 2

## The LinkPICS



# The LinkPICS architecture





## Extraction Step

### Ave mais idosa do Havaí sobrevive a tsunami e volta para cuidar de filhote

Conhecida e identificada, Wisdom tem cerca de 60 anos de idade. Onda cobriu até 60% de ilha na região e matou milhares de aves.

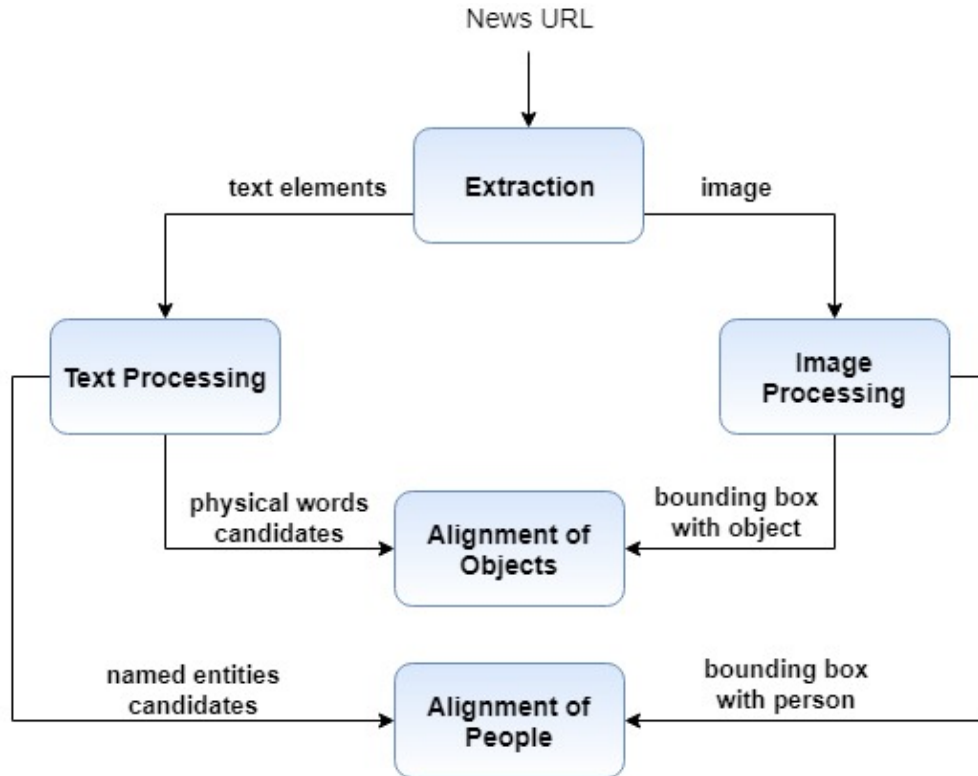
Do Globo Natureza, em São Paulo



- Textual elements
  - Title of the article
  - Text of the article
  - Image Caption
- Image elements
  - Image associated with the text



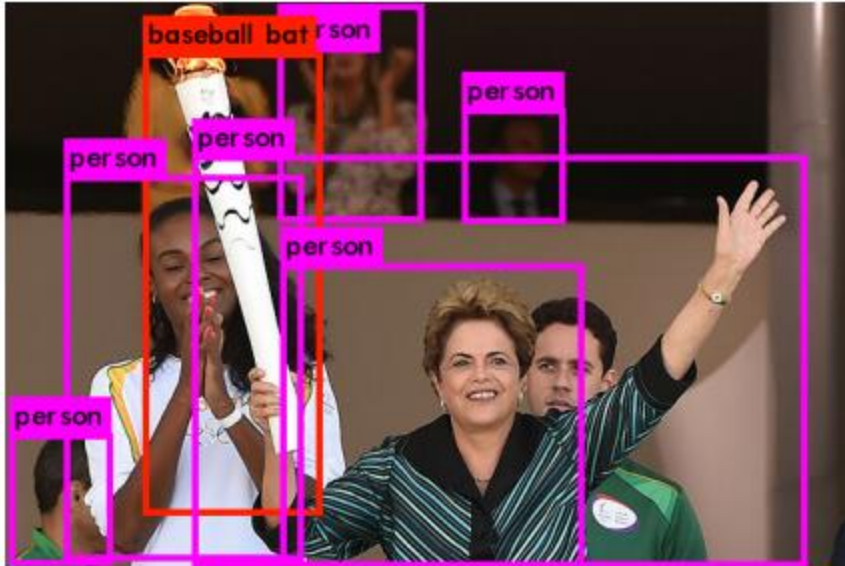
# The LinkPICS architecture







## Image processing step— Object Detector

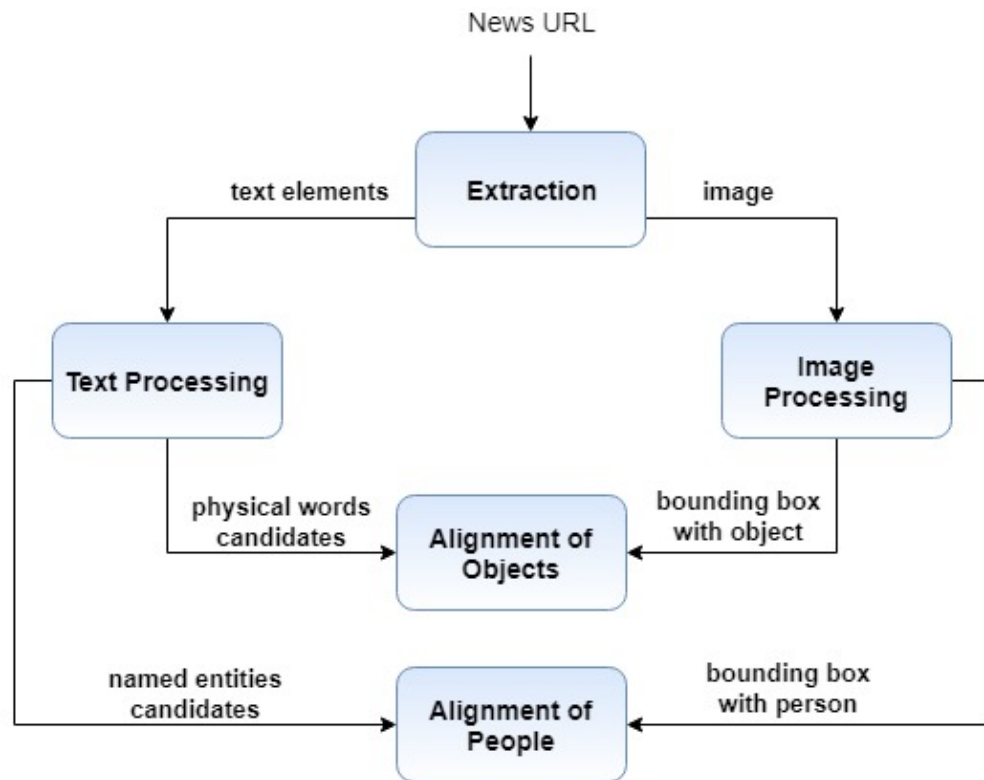


## Resources

- Object detector: YOLO [21]



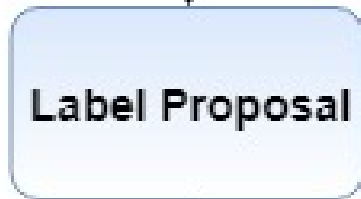
# The LinkPICS architecture





## Alignment of objects (Image Processing step)

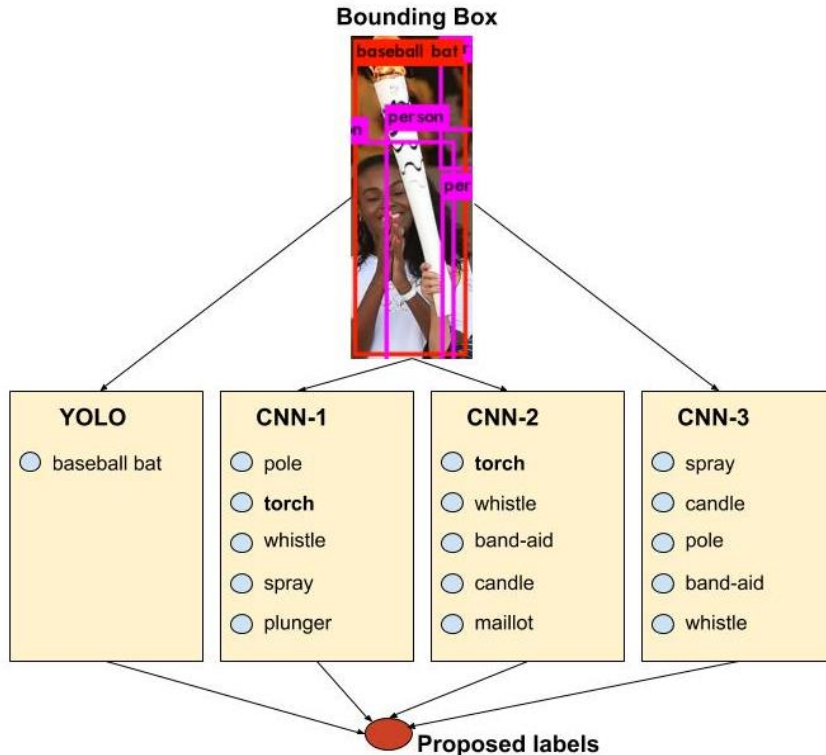
bounding box with object



Proposal labels



# Alignment of objects (Label Proposal)

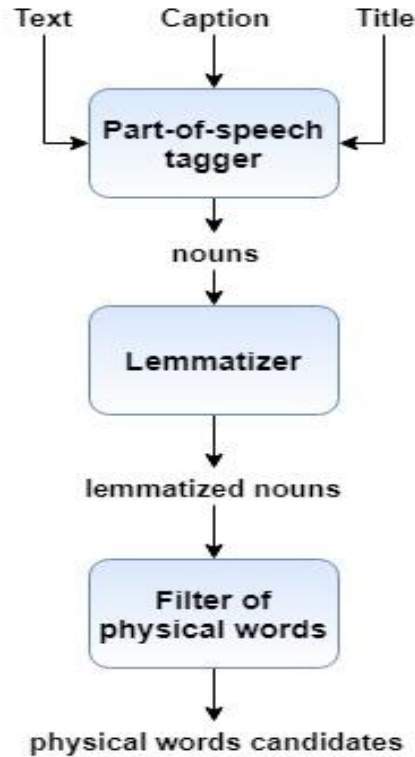


## Resources

- Word Embeddings: MUSE [4]
- Label proposal CNNs:
  - Darknet and Extraction<sup>1</sup>
  - DenseNet [10]



## Alignment of objects (Text Processing step)



## Resources

- Part-of-speech tagger: Tree Tagger [22]
- Lemmatizer: Tree Tagger [22]
- Filter of physical words: OpenWordNet PT [18]



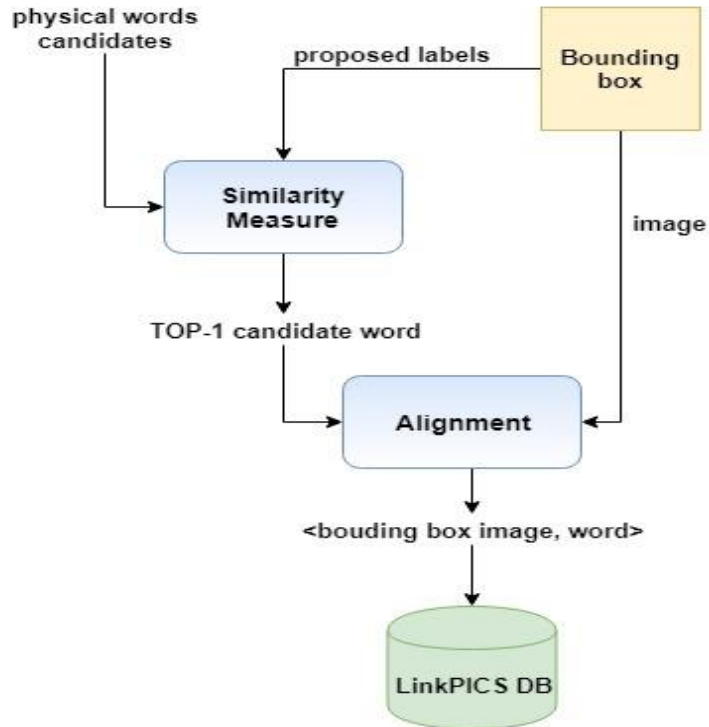
## Alignment of objects (Text Processing step)

**Example of Physical words candidates **output**:**

- torch, volleyball, games, champion ...

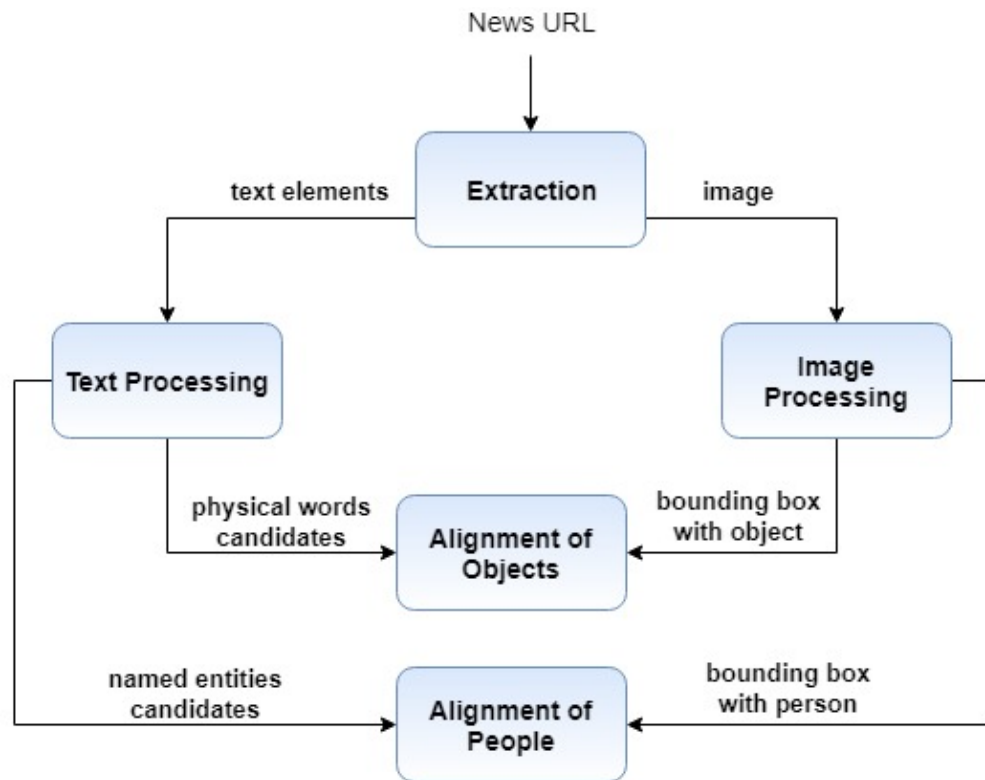


## Alignment of objects ( Alignment step)





# The LinkPICS architecture

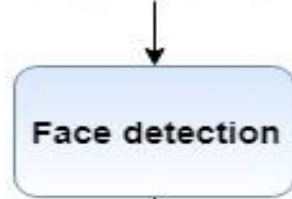






## Alignment of Person (Image Processing step)

bounding box with person



Face bounding box



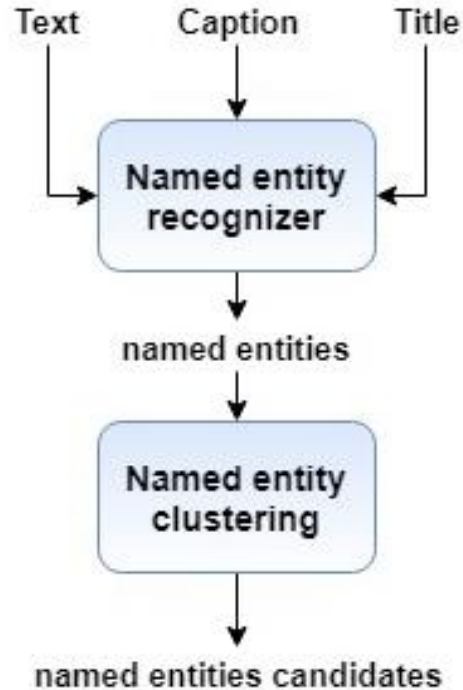
facial vector (128 elements)

## Resources

- Face detector: Haar Cascade [22]
- Generation of facial vector: FaceNet [23]



## Alignment of person (Text Processing phase)



## Resources

- Named entity recognizer: Polyglot [1]



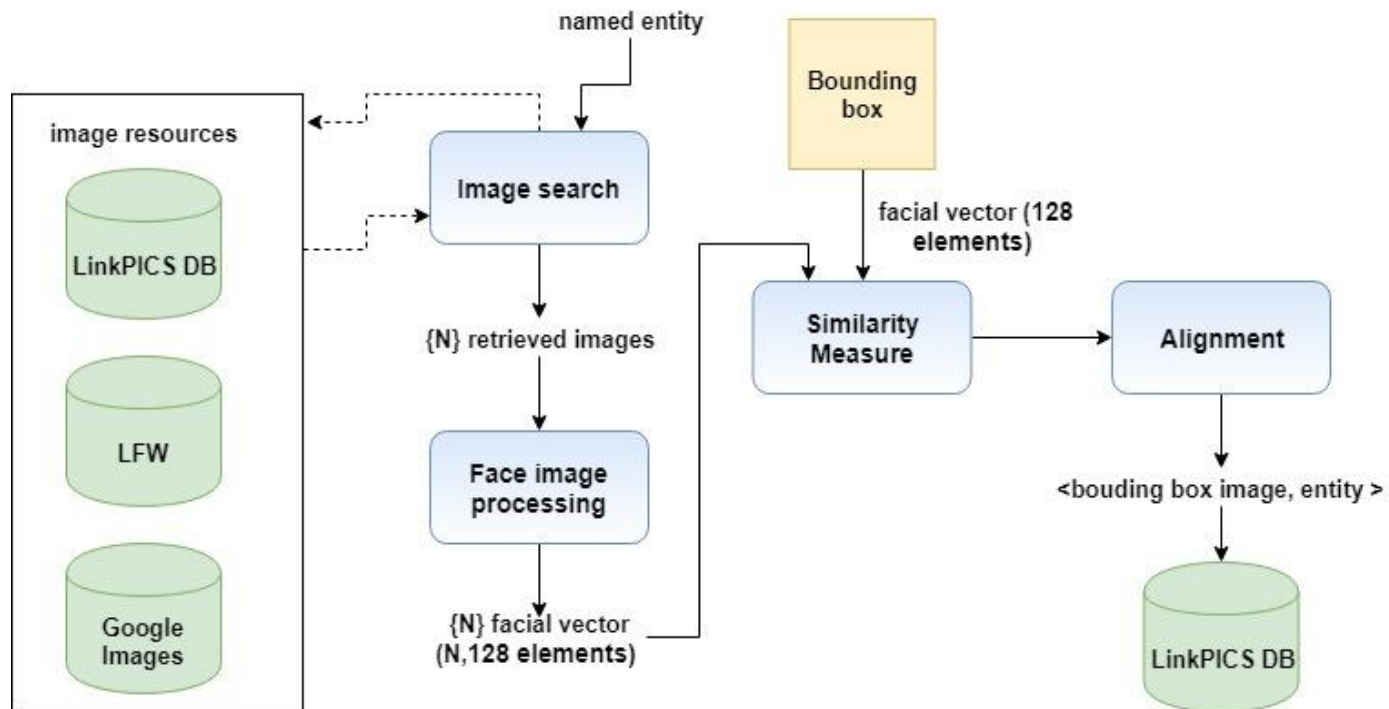
## Alignment of person (Text Processing phase)

**Example of Named entities candidates **output**:**

- {Dilma, Dilma Rousseff}, {Ayrton, Ayrton Senna},



## Alignment of Person ( Alignment step)



# 3

## Experiments and results



## Resources

- G1 corpus with 200 Portuguese News.

<sup>1</sup> Available in: [http:// g1.globo.com](http://g1.globo.com)

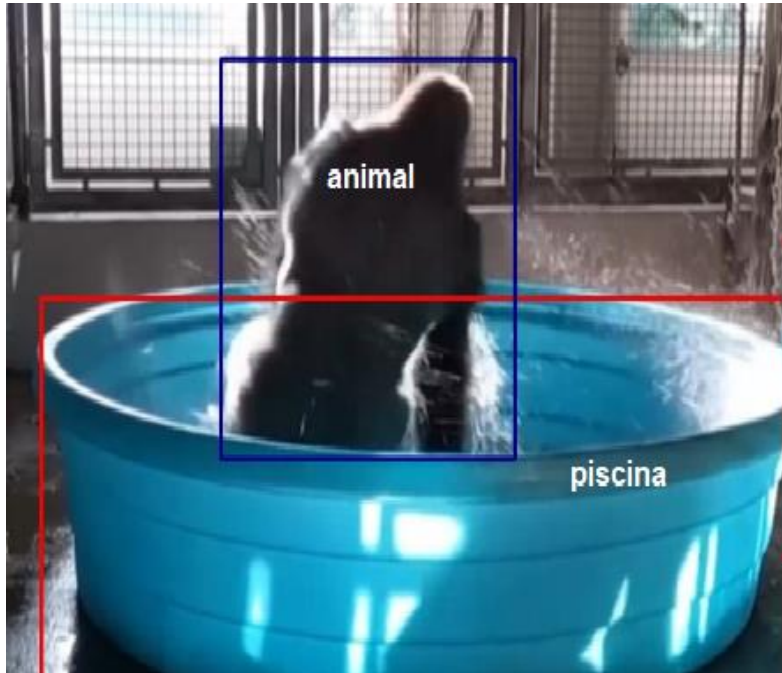


## Evaluation

	<b>Aligner</b>	<b>All</b>	<b>1 BB</b>	<b>2 or more BBs</b>
<b>People</b>	LinkPICS	95.52%	97.04%	90.74%
<b>Objects</b>	LinkPICS	73.39%	73.02%	75.86%



## Examples of LinkPICS alignment





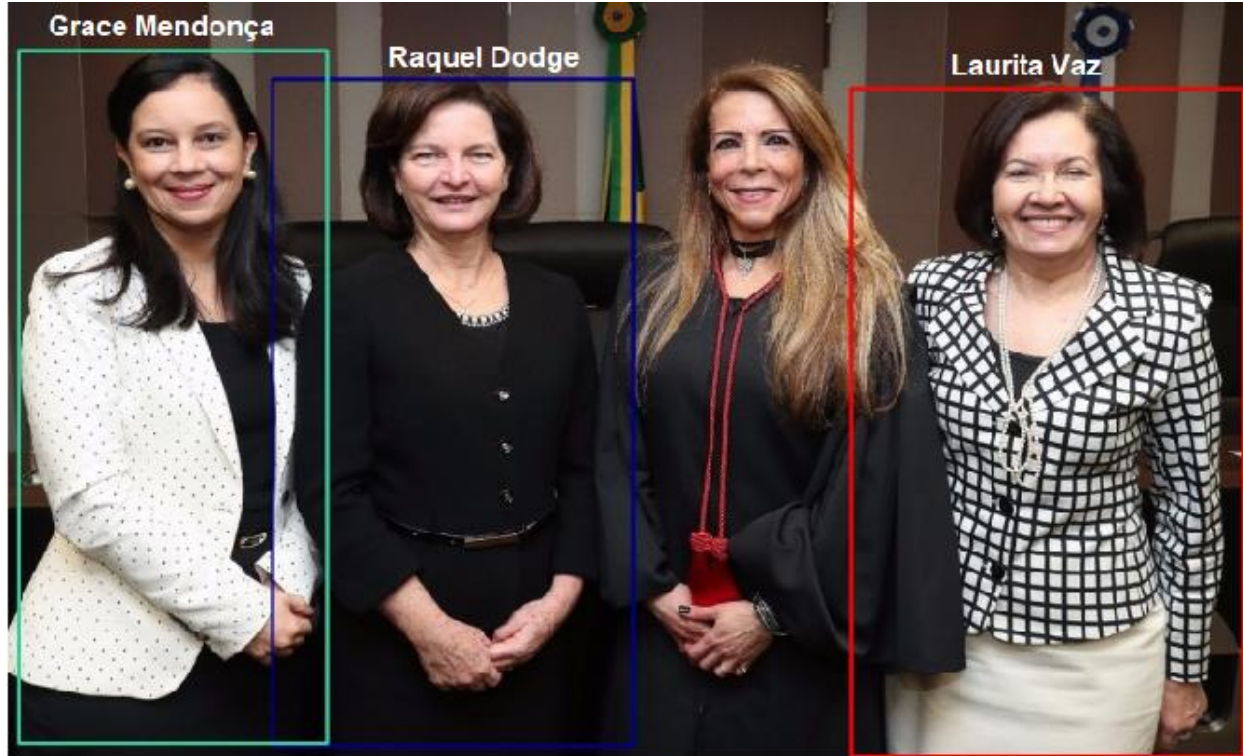


## Examples of LinkPICS alignment



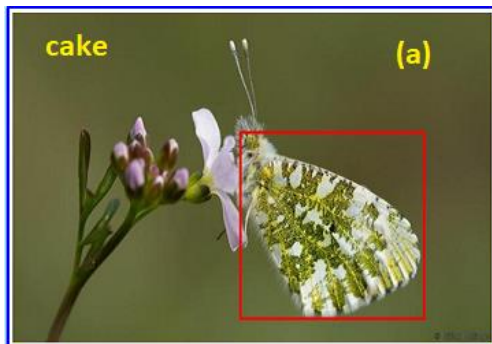


## Examples of LinkPICS alignment





# LinkPICS vs YOLO



# 4

## Conclusions and Future work



## Contributions

1. LinkPICS text-image aligner
2. G1 aligned corpus
3. LinkPICS database for visual dictionary
  - ▷ <named entity, face>
  - ▷ <physical word, object>



## Possible extensions

- Release LinkPICS for the research community.
- Output multiword expressions.
- N: 1 alignment ( several words to 1 bounding box.
- Alignment of landscapes (e.g., trees, mountains).
- Alignment of hard objects (e.g, drugs).



# THANKS!

**Any questions?**  
You can find me at  
[wellingtonveltroni@gmail.com](mailto:wellingtonveltroni@gmail.com)