

# CLASSIFICATION OF NOT SUITABLE FOR WORK IMAGES: A DEEP LEARNING APPROACH FOR ARQUIVO.PT

Daniel Bicho (a,b)

Artur Ferreira (b,c)

Nuno Datia (b,d)

- (a) Arquivo.pt
- (b) ISEL, Instituto Superior de Engenharia de Lisboa
- (c) Instituto de Telecomunicações
- (d) NovaLincs, FCT, Universidade Nova de Lisboa



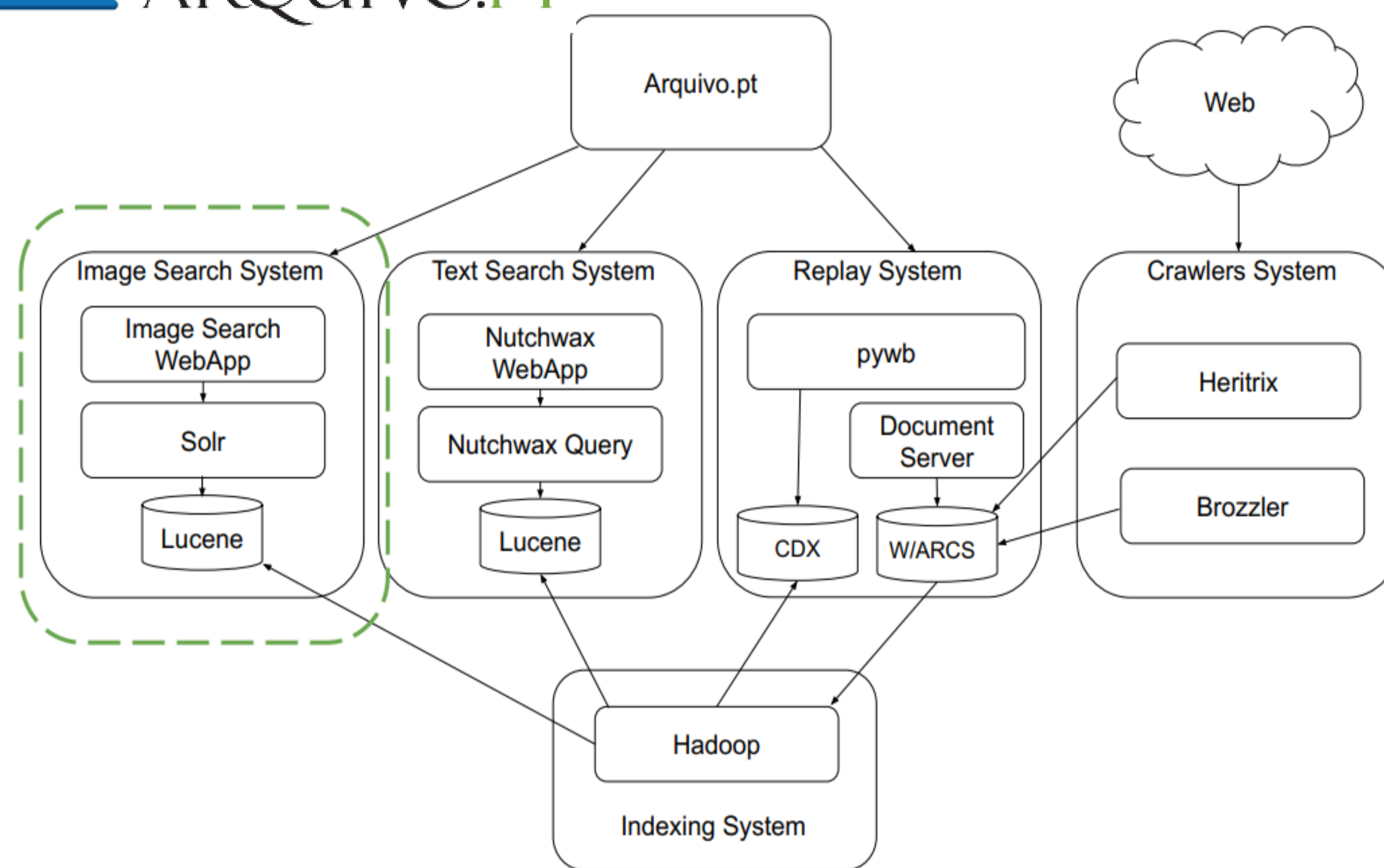
## WEB ARCHIVING

- Aims to preserve Web contents for historical purposes
- Web crawlers search the web
- The retrieved web contents are stored as archives
- Arquivo.pt preserves the .pt top-level domain



NUMBER OF RESOURCES MIME-TYPES COLLECTED ON ARQUIVO.PT  
LAST BROAD DOMAIN CRAWL.

% Amount	Number of URL	mime-types
79.60	237 966 251	text/html
10.03	29 997 689	image/jpeg
2.45	7 328 179	image/png
0.77	2 305 834	application/pdf
0.76	2 271 770	application/javascript
0.72	2 145 481	text/xml
0.62	1 869 902	application/rss+xml
0.62	1 861 165	application/json
0.60	1 820 236	image/gif
0.58	1 783 630	text/css
3.25	1 783 630	all others

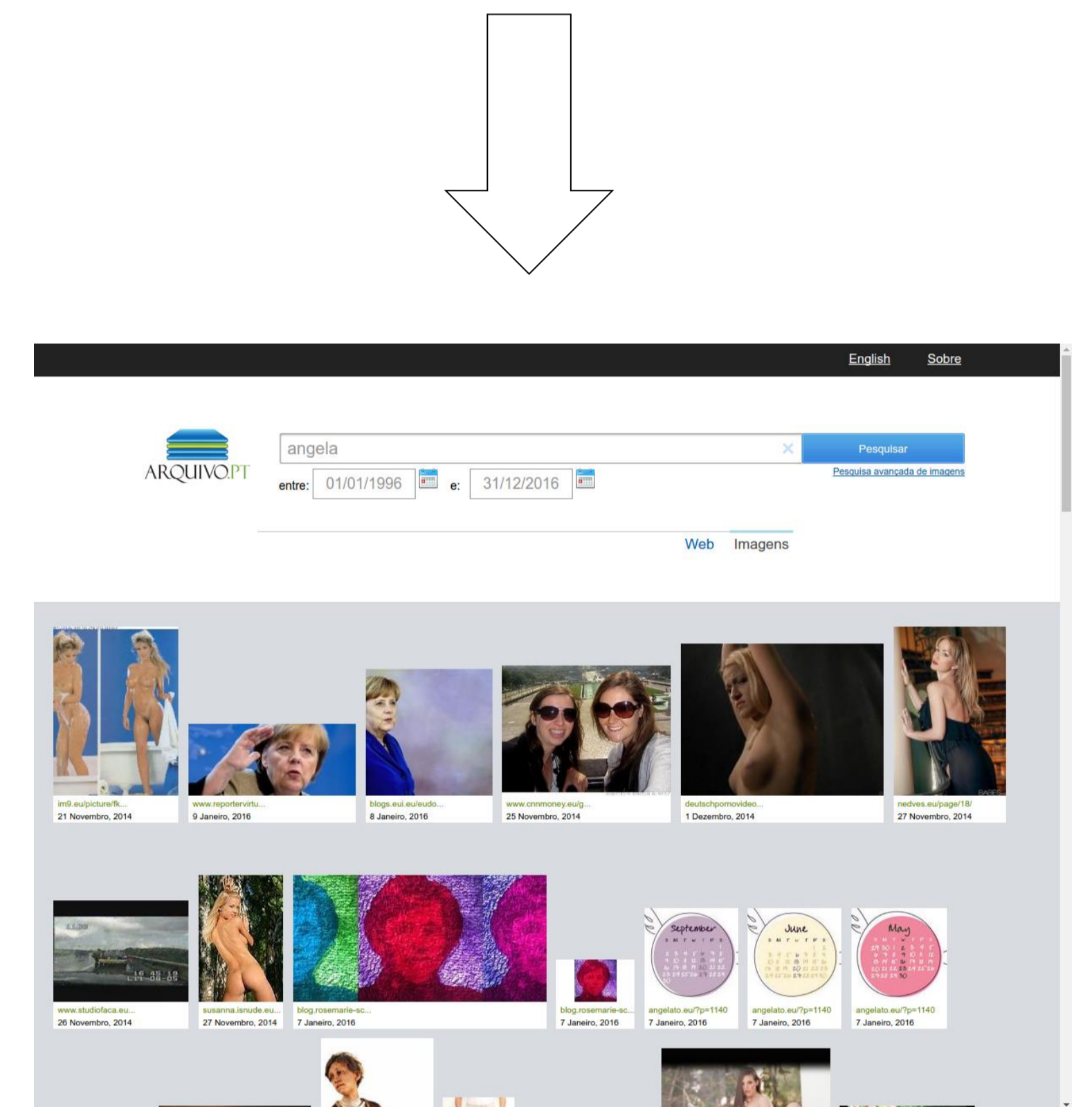


MIME-TYPES DISTRIBUTION FOR IMAGE TYPE CONTENTS.

% Amount	Number of URL	mime-types
75.59	30 145 538	image/jpeg
18.37	7 328 179	image/png
4.56	1 820 236	image/gif
0.98	389 839	image/svg+xml
0.34	135 720	image/x-icon
0.09	38 577	image/pjpeg
0.06	22 717	image/bmp

## THE PROBLEM

- Some images are Not Suitable For Work (NSFW)
- Nudity and pornographic contents
  - Harmful for many users
  - Websites may be hacked before the crawling action
- Retrieved results with query term "angela" on arquivo.pt



## PROPOSED SOLUTION (IN 4 STEPS)

### 1) Build a dataset

- Acquire many images, with spatial resolution above 150x150 pixels
- With the existent Beta images indexes
- With the text search API

SUMMARY OF THE COLLECTED DATASET.

	SFW	NSFW	Total
Labeled Images	9 382	8 273	17 655
Non-Labeled Images	-	-	18 626

### 2) Evaluation of DNN Models

- OpenNSFW** and **NSFWsqueezeNet** models
  - Binary classification task
  - Loss function: cross-entropy
  - Optimizer: Stochastic Gradient Descent

	NSFW'	SFW'
NSFW	7706	567
SFW	652	8730

93% accuracy

	NSFW'	SFW'
NSFW	3697	4576
SFW	333	9049

72% accuracy

### 3) Fine Tuning of the Models

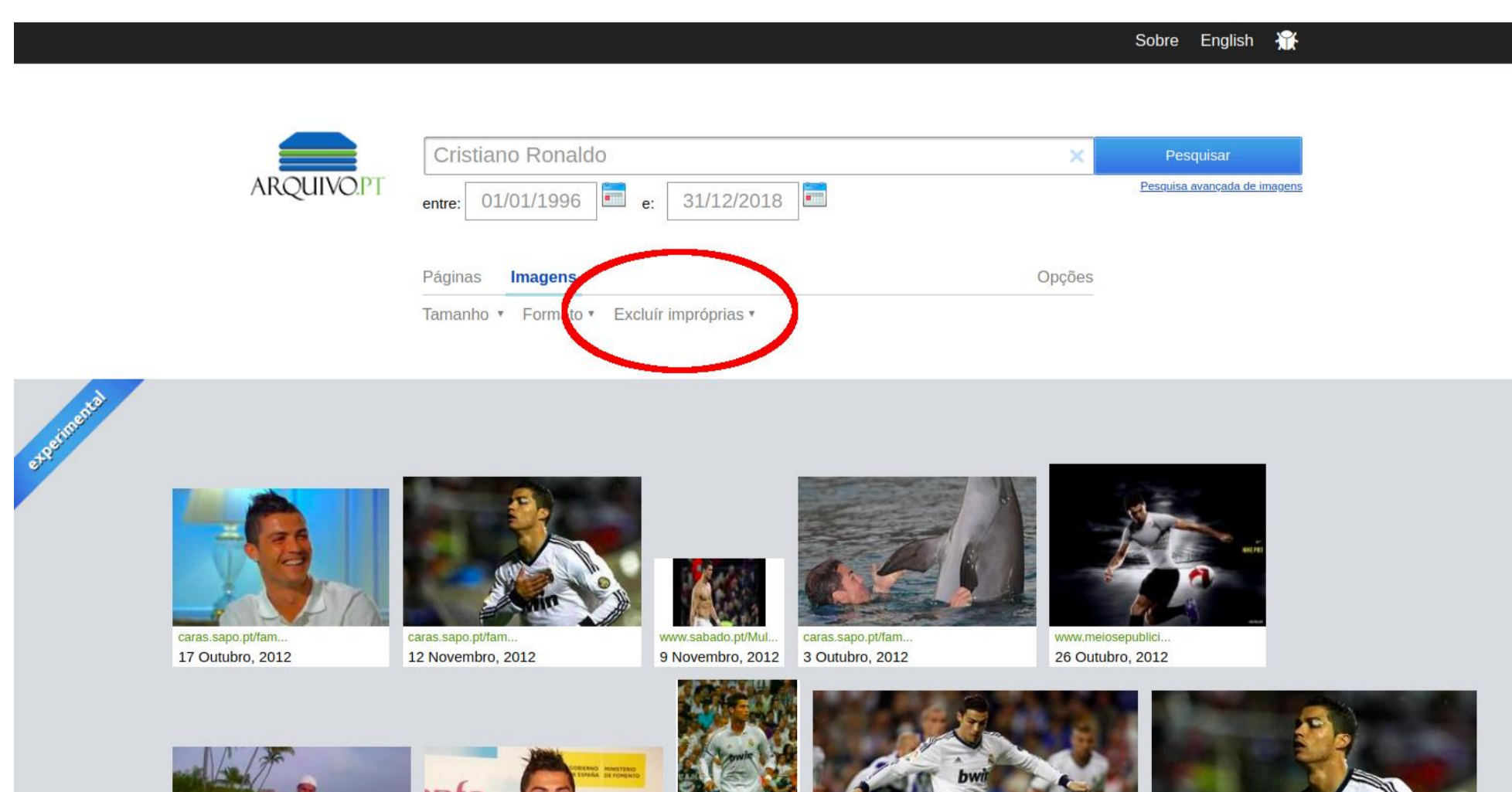
- Retrain all the network layers
- Starting point - the network weights from a pre-trained model
- Improved accuracies:
  - OpenNSFW – 94%
  - NSFWsqueezeNet – 89%

### 4) Integration into Arquivo.pt Image Search System

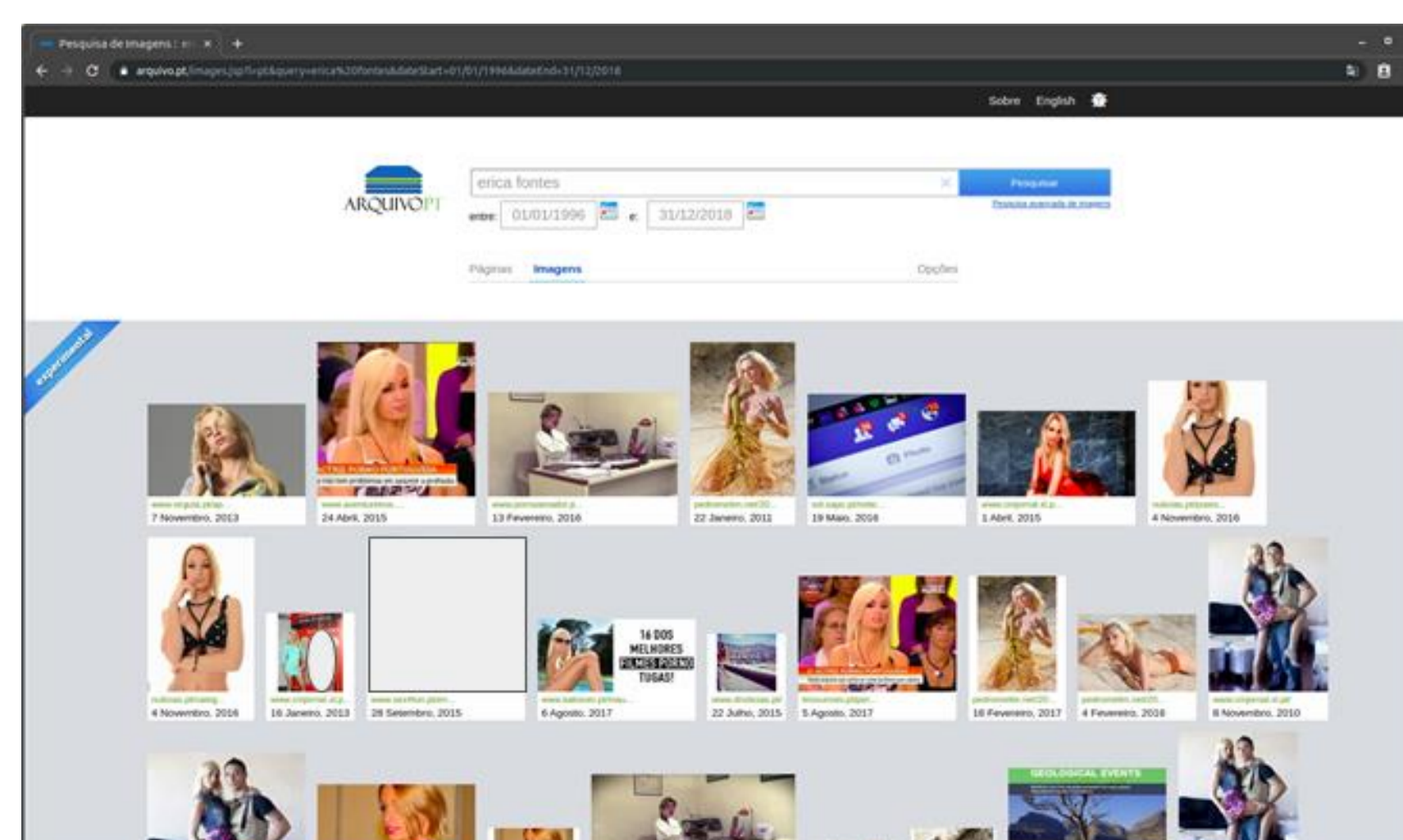
- a) Image classification interface integration, using as query term 'Cristiano Ronaldo'. The red ellipse highlights the NSFW filter choice by the user
- b) Query term 'Erica Fontes', with the NSFW classifier. The gray rectangular box highlights NSFW contents which were misclassified
- c) Query term 'Erica Fontes', without the NSFW classifier. The gray rectangular boxes highlight (and hide) NSFW contents

## REMARKS

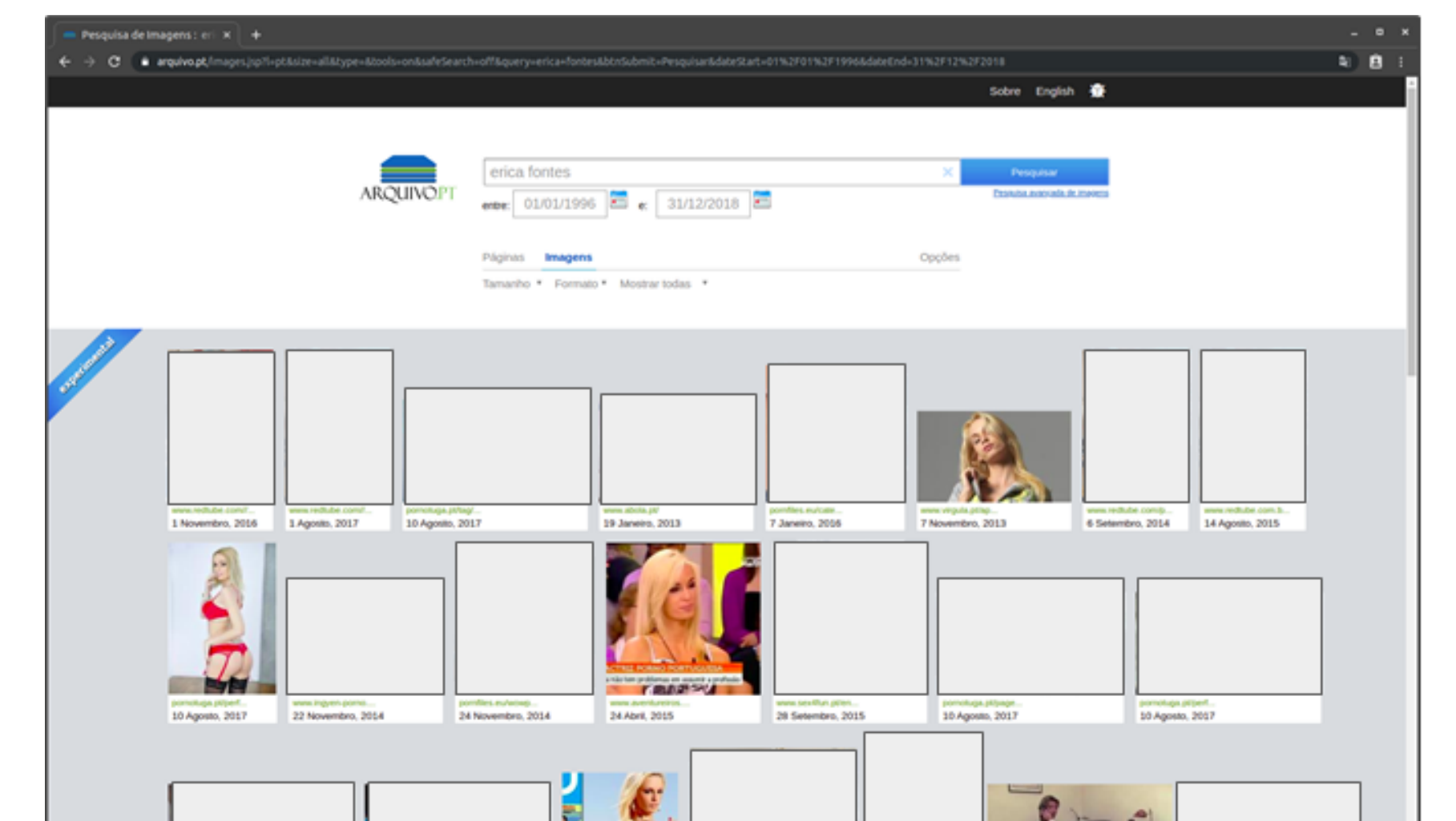
- A solution that automatically identifies not suitable for work images
- Integrated into Arquivo.pt infrastructure
- The model accuracy can be improved
  - with a larger dataset
  - considering a multi-class approach



a)



b)



c)