

conta-me
histórias



conta-me histórias

temporal summarization and the Portuguese
web archive

Ricardo Campos, Arian Pasquali, Vítor Mangaravite,
Alípio Jorge, Adam Jatowt

TPDL 2018



about me

arian pasquali

Researcher

University of Porto and LIAAD - INESC TEC



challenges

- . information overload
- . credibility crisis
- . fake news and post-truth era
- . filter bubbles
- . content based machine learning models (e.g. fact-checkers)
- . more than ever web archives are crucial resources



importance of web archives

the importance of preserving the history of the web

~ 2%

of the websites
disappear
every week *

~ 8%

of urls disappear
every year **

* Frank McCown, Catherine C. Marshall, and Michael L. Nelson. 2009. Why web sites are lost. *Commun. ACM* 52, 11 (November 2009), 141-145.

** Legal Information Archive, The chesapeake project report. 2008

2018 Arquivo.pt announces a free text search API

Challenge Arquivo.pt Prizes

- Use public API to create innovative ways to explore the web archive

<https://sobre.arquivo.pt/en/collaborate/submissions-for-the-arquivo-pt-prizes-2018/>

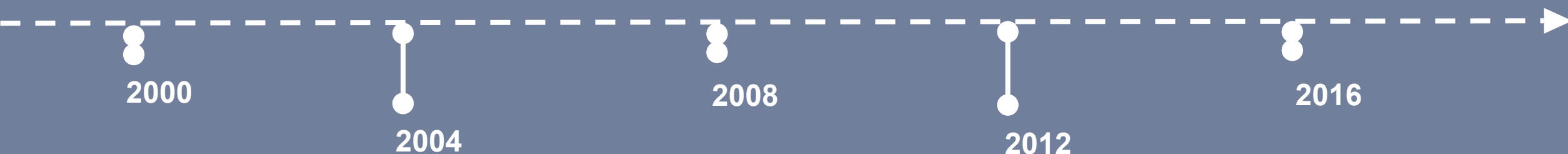


The screenshot displays the Arquivo.pt search interface. At the top, the logo consists of a blue icon of stacked books followed by the text "ARQUIVO.PT" in green. Below the logo is a search bar containing the text "troika em Portugal" and a blue "Pesquisar" button. To the right of the button is a link for "Pesquisa avançada". Below the search bar, there are date selection fields: "entre: 01/01/1996" and "e: 31/12/2017", each with a small calendar icon. A horizontal line separates the search area from the results. On the right side, it says "Resultados 1 a 10 de 23.776.571". The first result is titled "Livro Os 10 erros da Troika em Portugal - Figueiras (Porto) ..." with a date of "15 Abril, 2015" and a link to "Listar versões". The description mentions "Livro Os 10 erros da Troika em Portugal - Figueiras (Porto) - Livros - Revistas" and includes a URL: "http://figueiras.olx.pt/livro-os-10-erros-da-troika-em-portugal-iid-463457489". The second result is titled "Louça quer saber para onde vai 1.ª tranche da 'troika' -" with a date of "31 Maio, 2011" and a link to "Listar versões". The description mentions "Louça quer saber para onde vai 1.ª tranche da 'troika' - Portugal - DN Última hora" and includes a URL: "http://www.dn.pt/inicio/portugal/interior.aspx?content_id=1866281". The third result is titled "\"Este programa está muito além do memorando\" da 'troika' -" with a date of "8 Maio, 2011" and a link to "Listar versões". The description mentions "\"Este programa está muito além do memorando\" da 'troika' - Portugal - DN Última hora" and includes a URL: "http://www.dn.pt/inicio/portugal/interior.aspx?content_id=1847261".

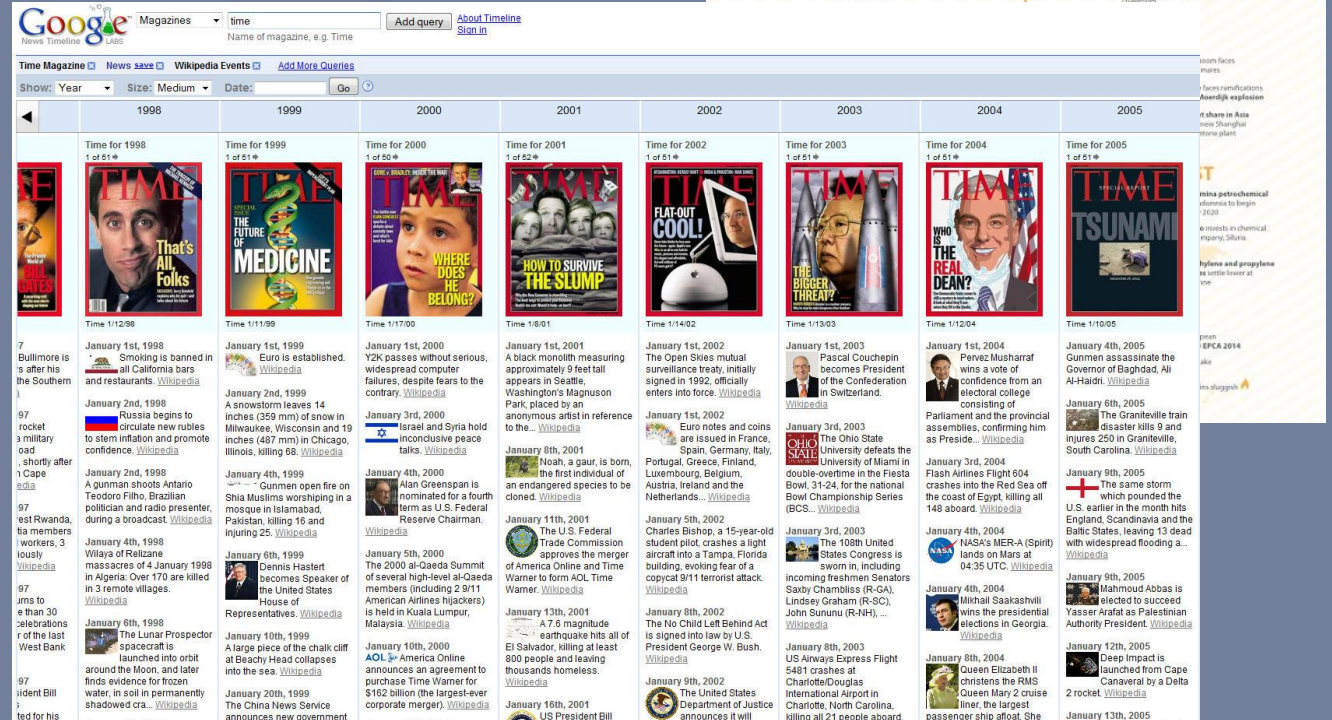
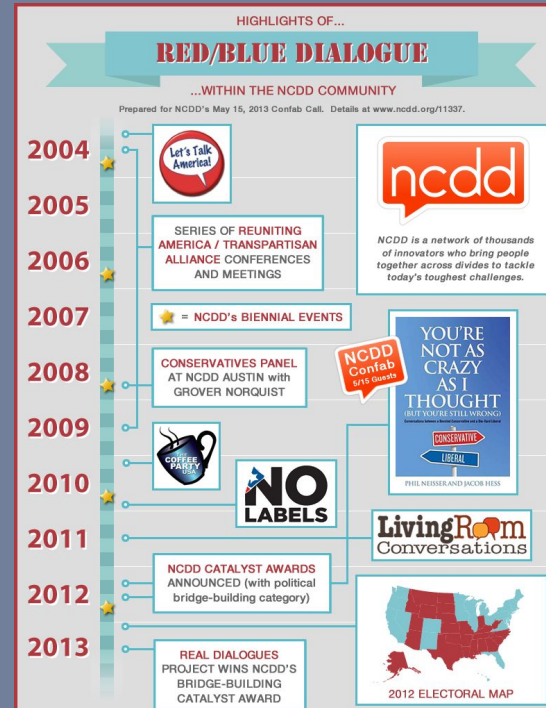
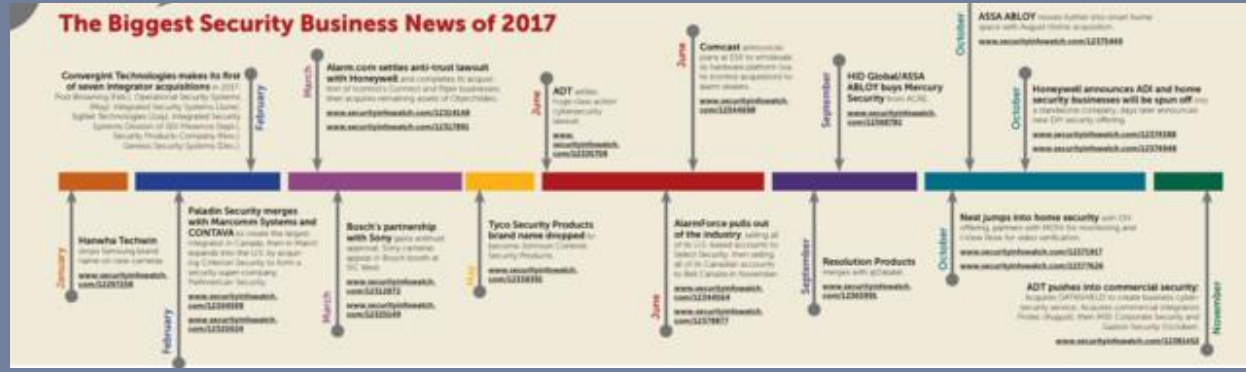
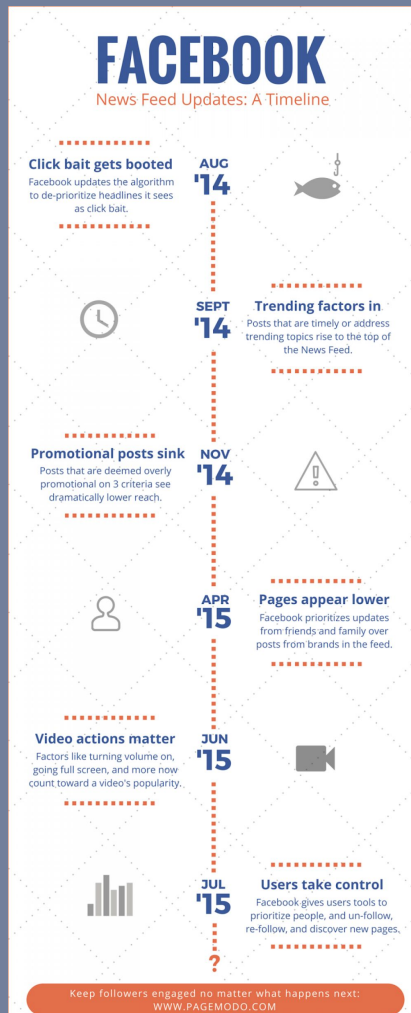
idea

what if we could generate a timeline about
any given topic automatically?

can we build a left-wing versus right-wing
narrative timeline?



timelines are a natural choice to explore long term stories

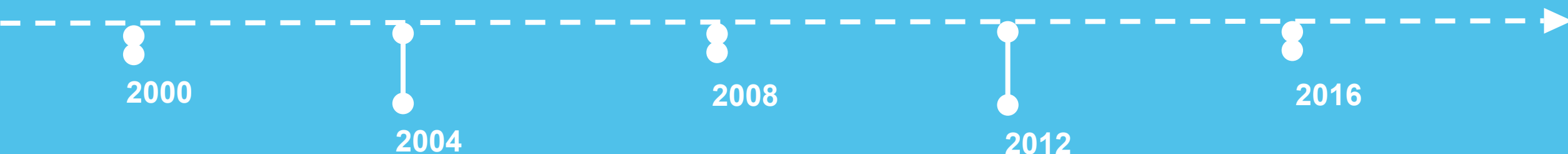


problem definition

given a search **query**
and a **timeframe**

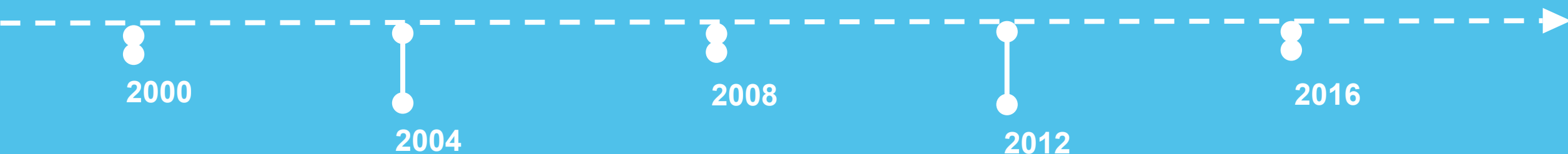


generate a timeline
summarization with relevant
dates and headlines



steps

1. query news articles using arquivo.pt API
2. identify relevant time frames
3. identify relevant headlines



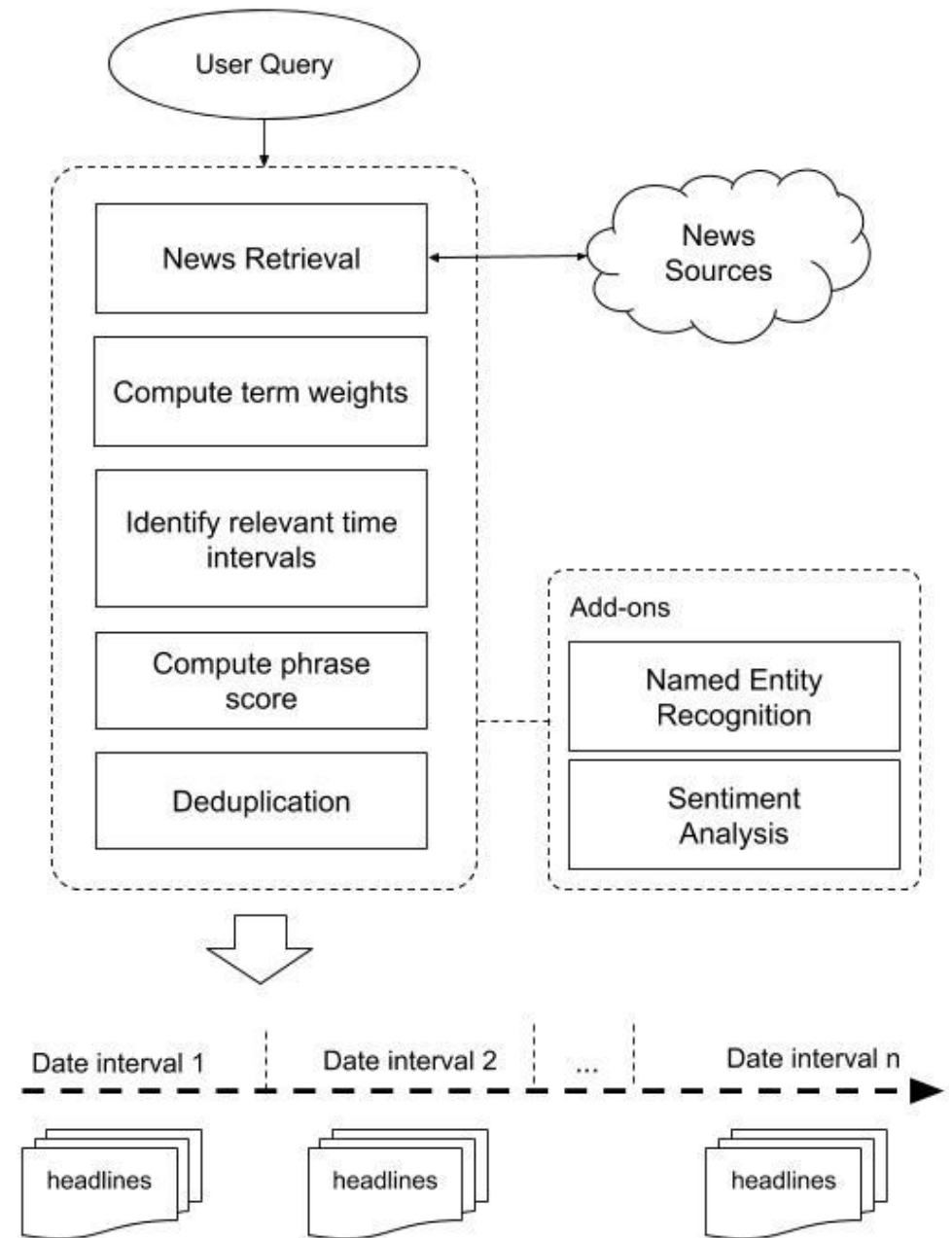
how to calculate headline relevance?

we apply principles of keyword extraction algorithms to calculate headlines relevance from a particular time frame



architecture

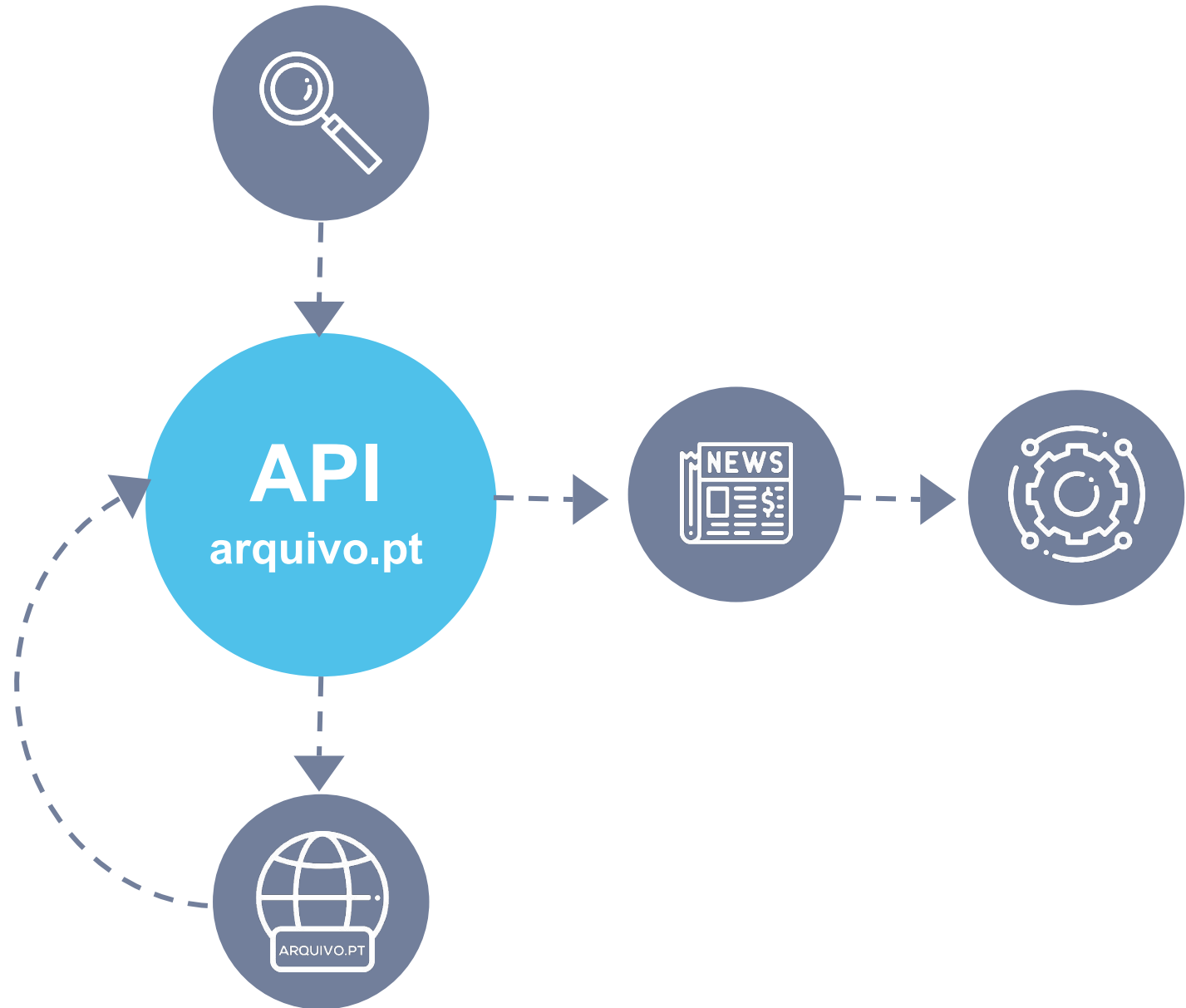
- given a search **query** and a **timeframe**
- apply principles of keyword extraction to compute relevant headlines from a particular time frame



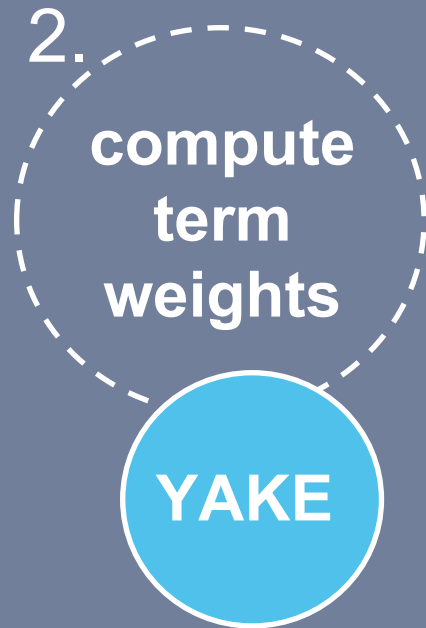
architecture



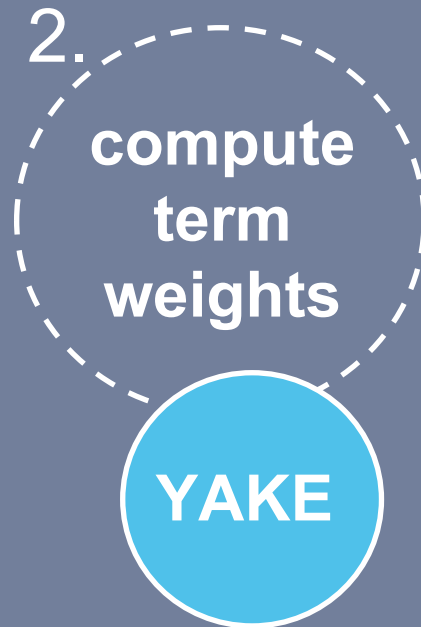
architecture



architecture



architecture



More information

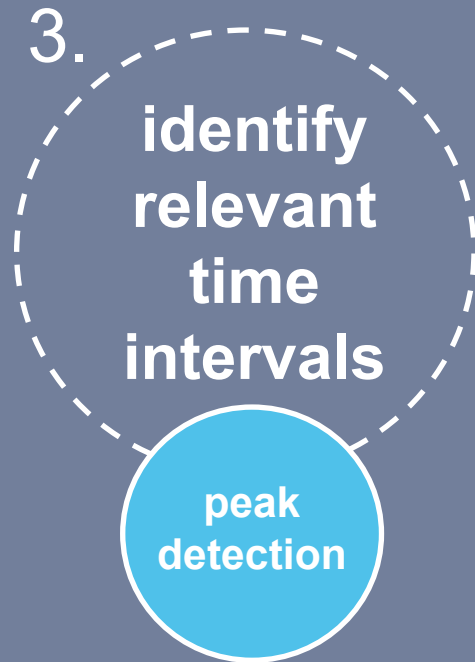
Campos, R., & Mangaravite, V., & Pasquali, A., & Jorge, A., & Nunes, C., & Jatowt, A. (2018). A Text Feature Based Automatic Keyword Extraction Method for Single Documents. Best Short Paper of ECIR'18 (<http://bit.ly/YakeDemoECIR2018>)

$$S(t) = \frac{TC_{Context} * TPos}{TC_{Case} + \frac{TFreq}{TC_{Context}} + \frac{TSent}{TC_{Context}}}$$

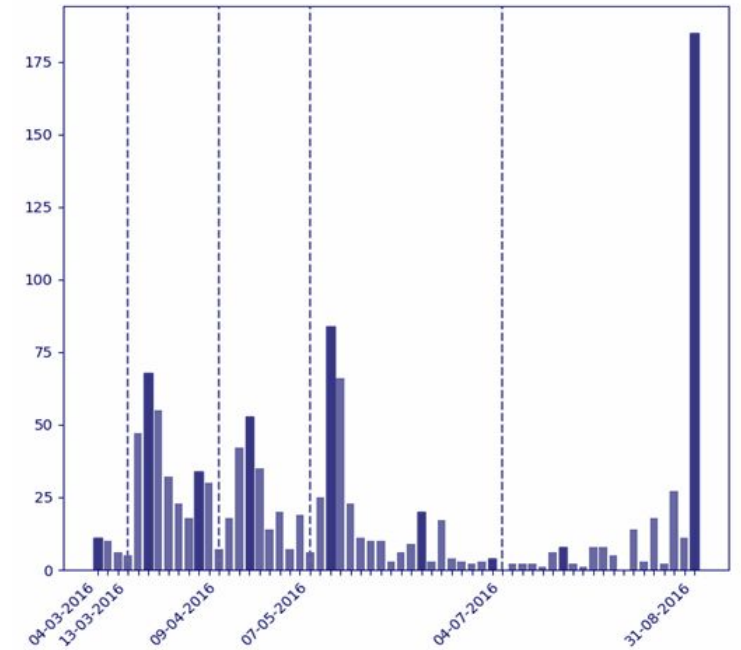
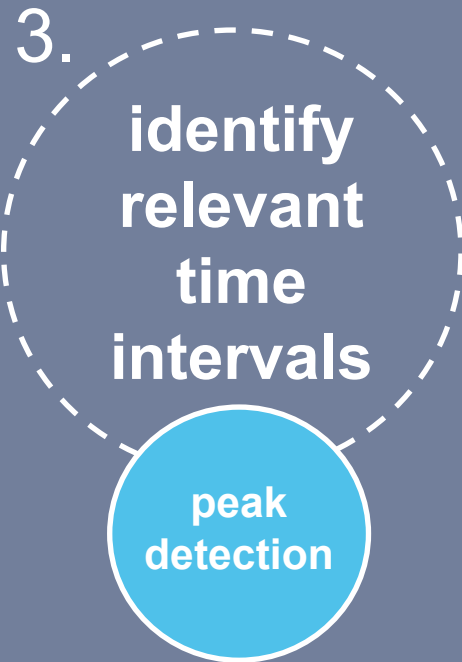
where:

- **TCase** - Casing
 - Reflects case aspect of a particular word
- **TPos** - Term position
 - Relevant terms tend to occur at the beginning of the document
- **TFreq** - Term frequency
 - Assumption that relevant words occur less
- **TSent** - term frequency on different headlines
 - How often a word appears within different headlines
- **TContent** - Term relatedness to context
 - Number of different terms that occur to the left (or right)
 - the more the number of different words around the the candidate word, more meaningless the word is likely to be

architecture



architecture



1. We aggregate count of news by date

2. We then detect peaks using scipy's relative extrema function

```
from scipy.signal import argrelextrema
```

```
x = np.array([2, 1, 2, 3, 2, 0, 1, 0])  
argrelextrema(x, np.greater, order=1)
```

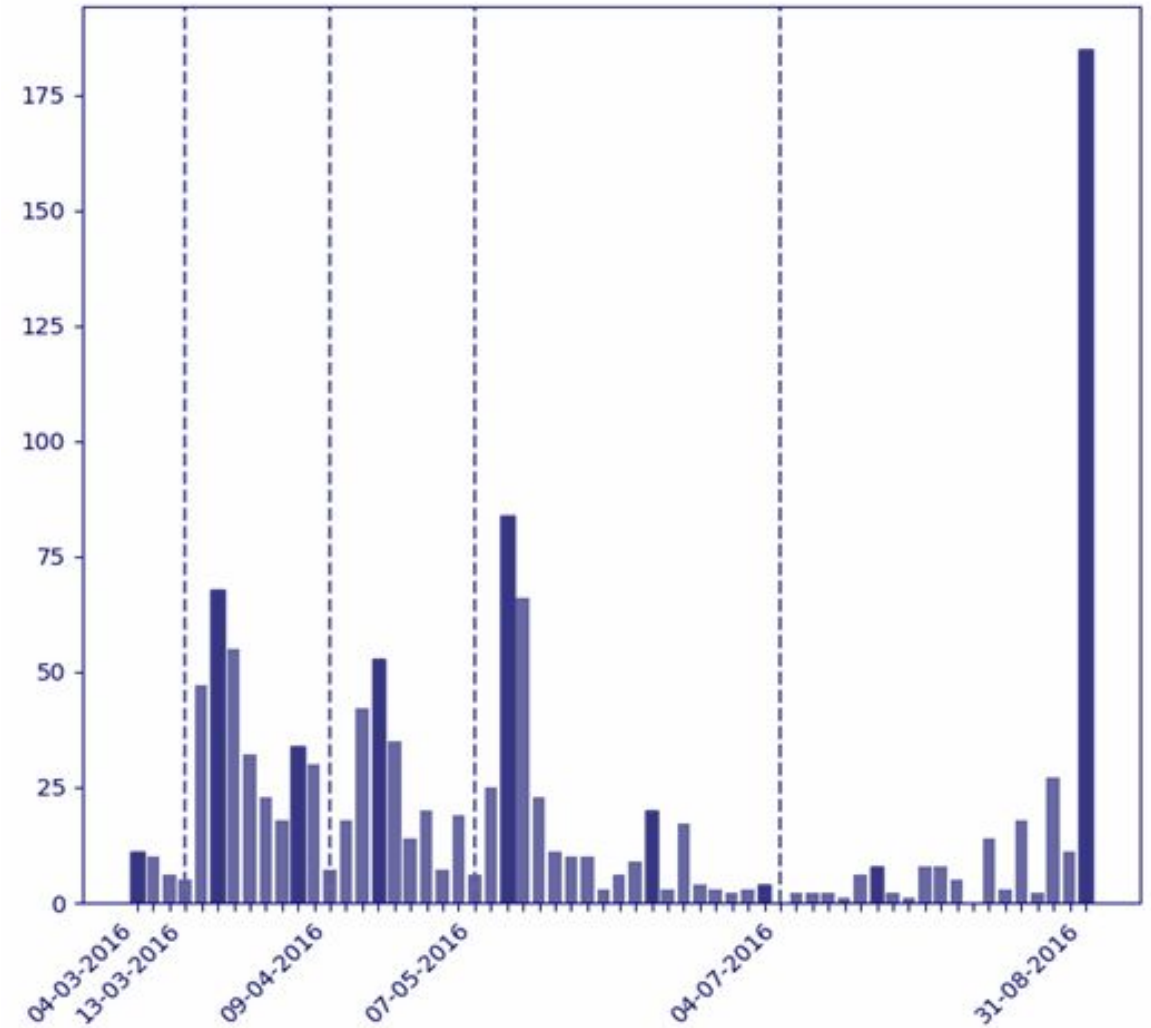
```
>> (array([3, 6]),)
```


architecture

3.

identify
relevant
time
intervals

peak
detection

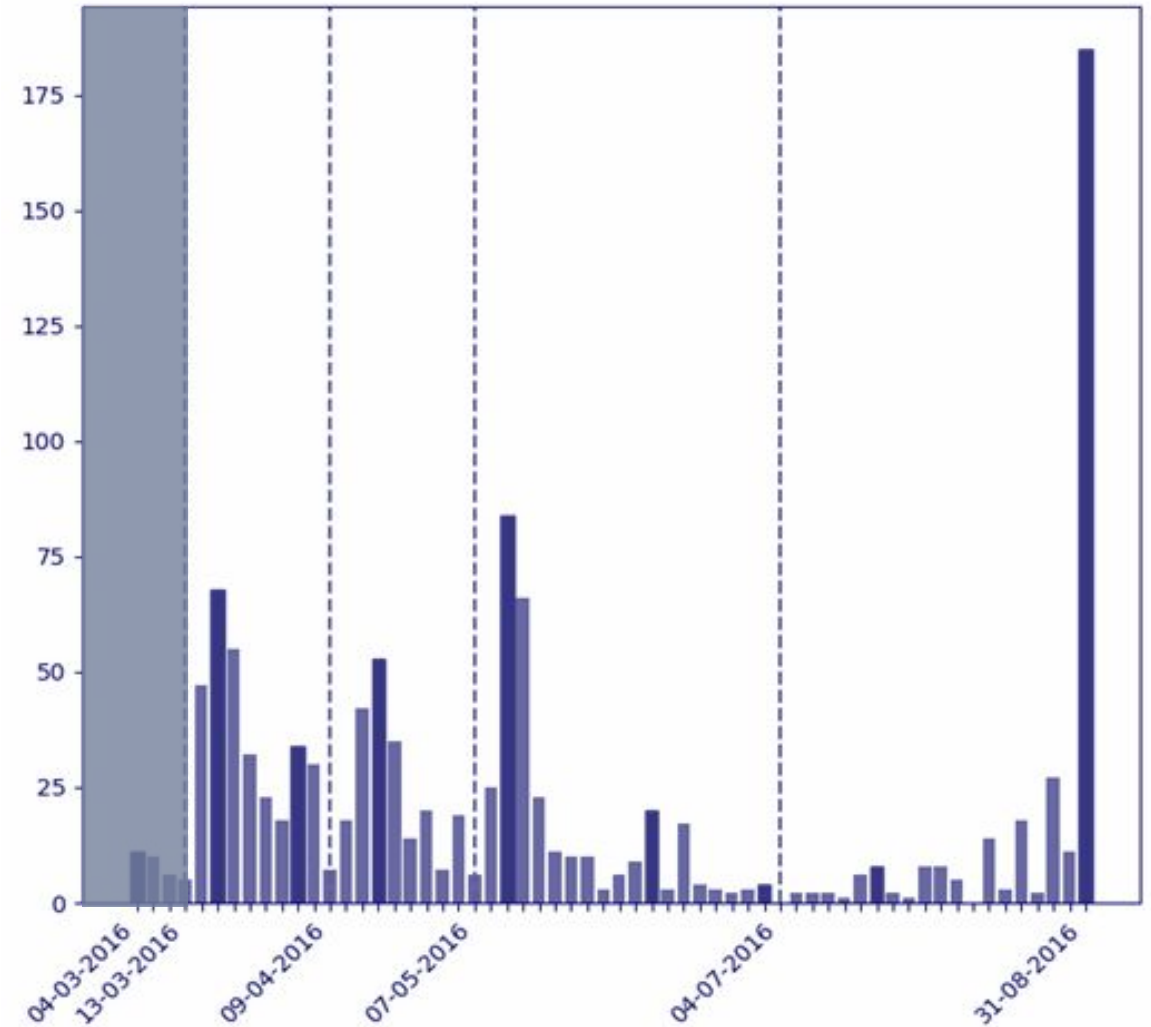


architecture

3.

identify
relevant
time
intervals

peak
detection

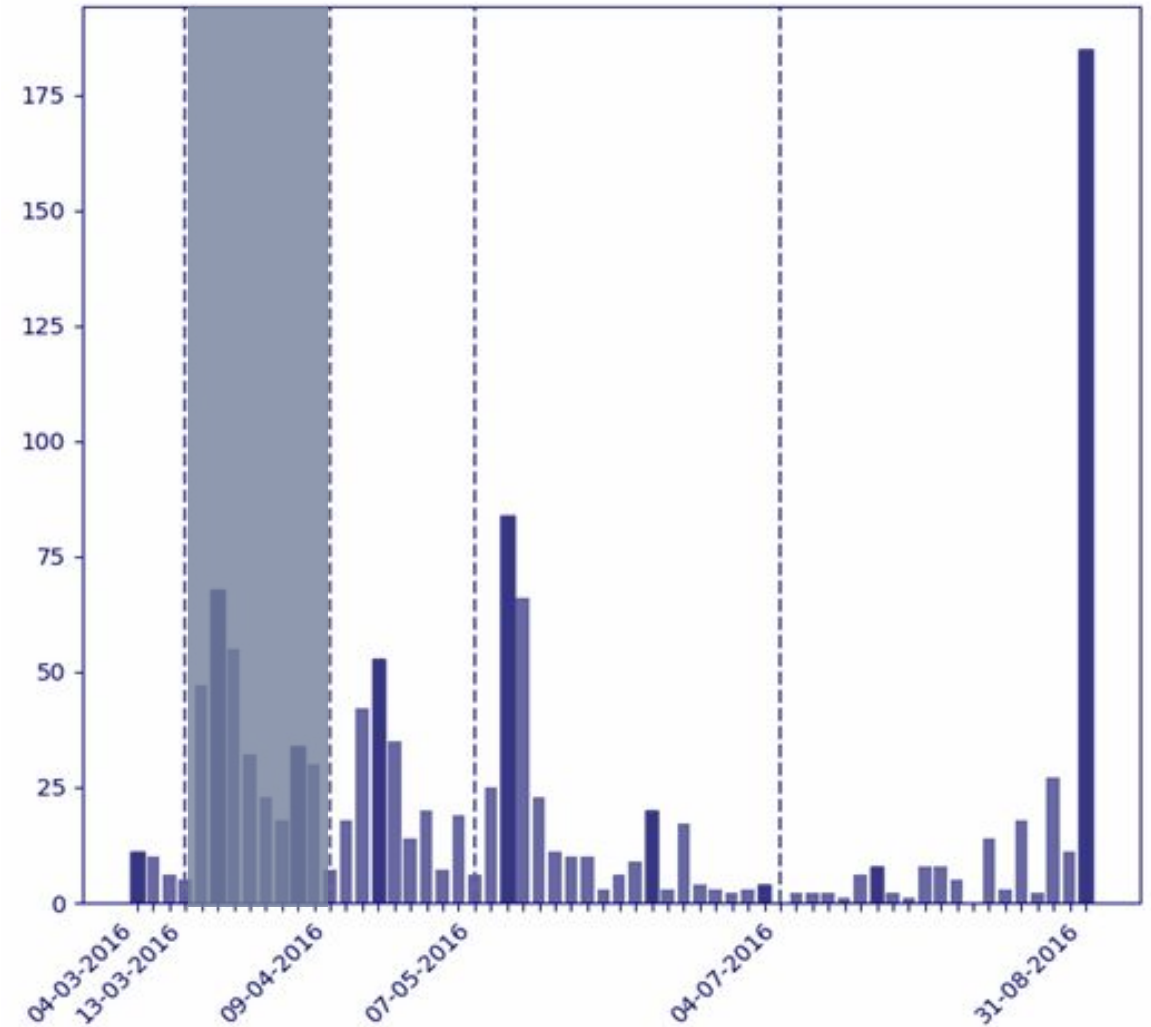


architecture

3.

identify
relevant
time
intervals

peak
detection

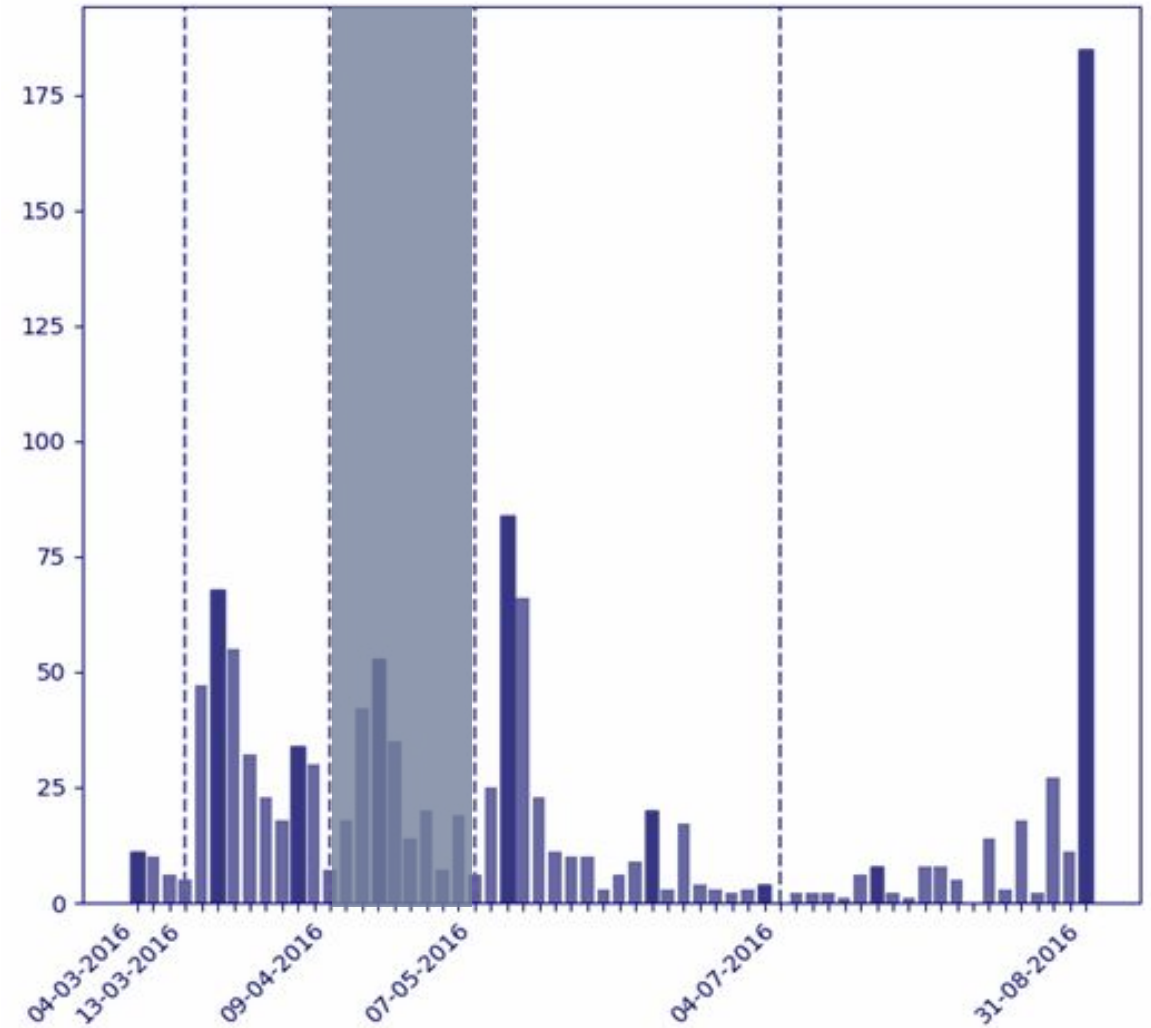


architecture

3.

identify
relevant
time
intervals

peak
detection

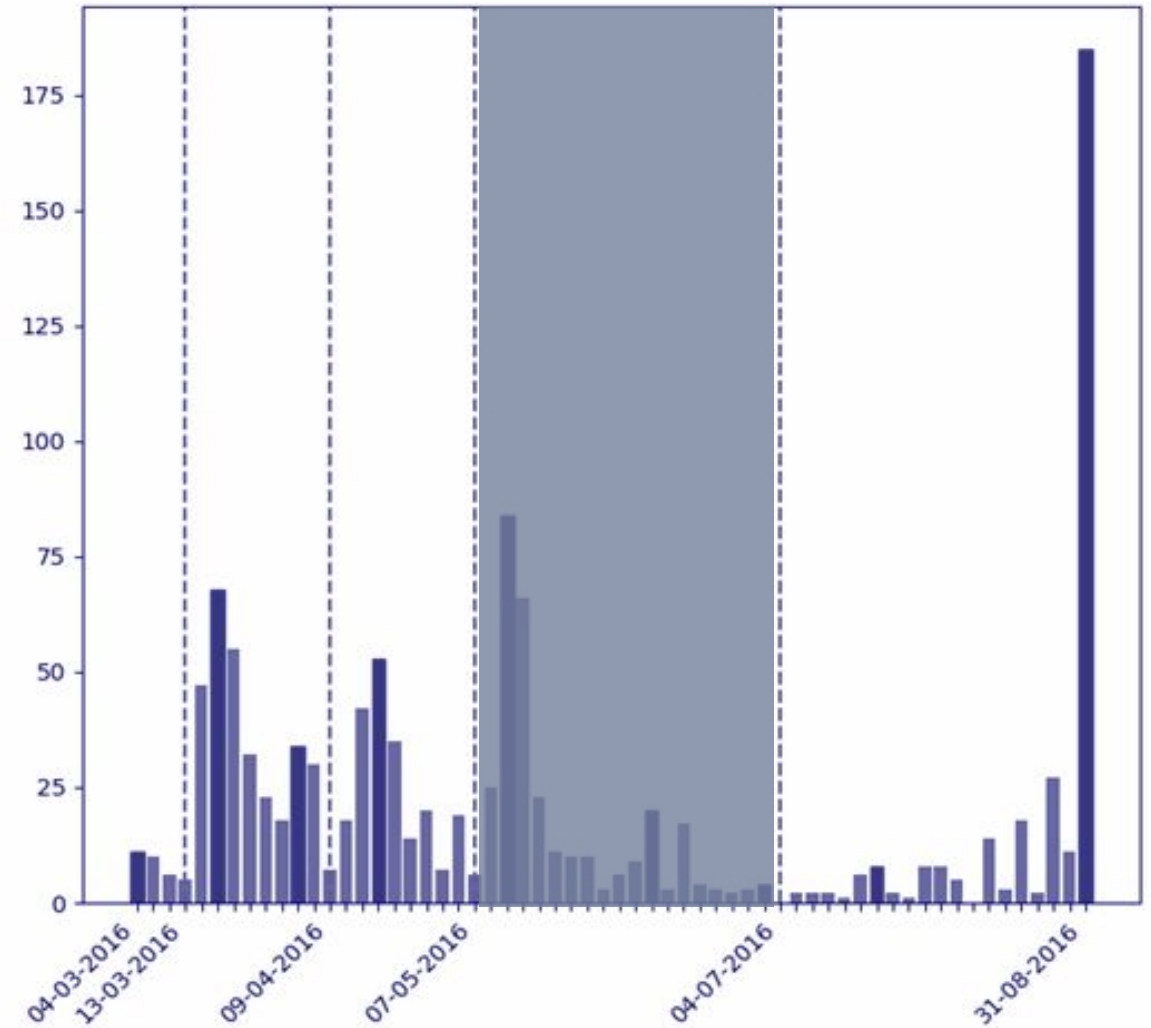


architecture

3.

identify
relevant
time
intervals

peak
detection

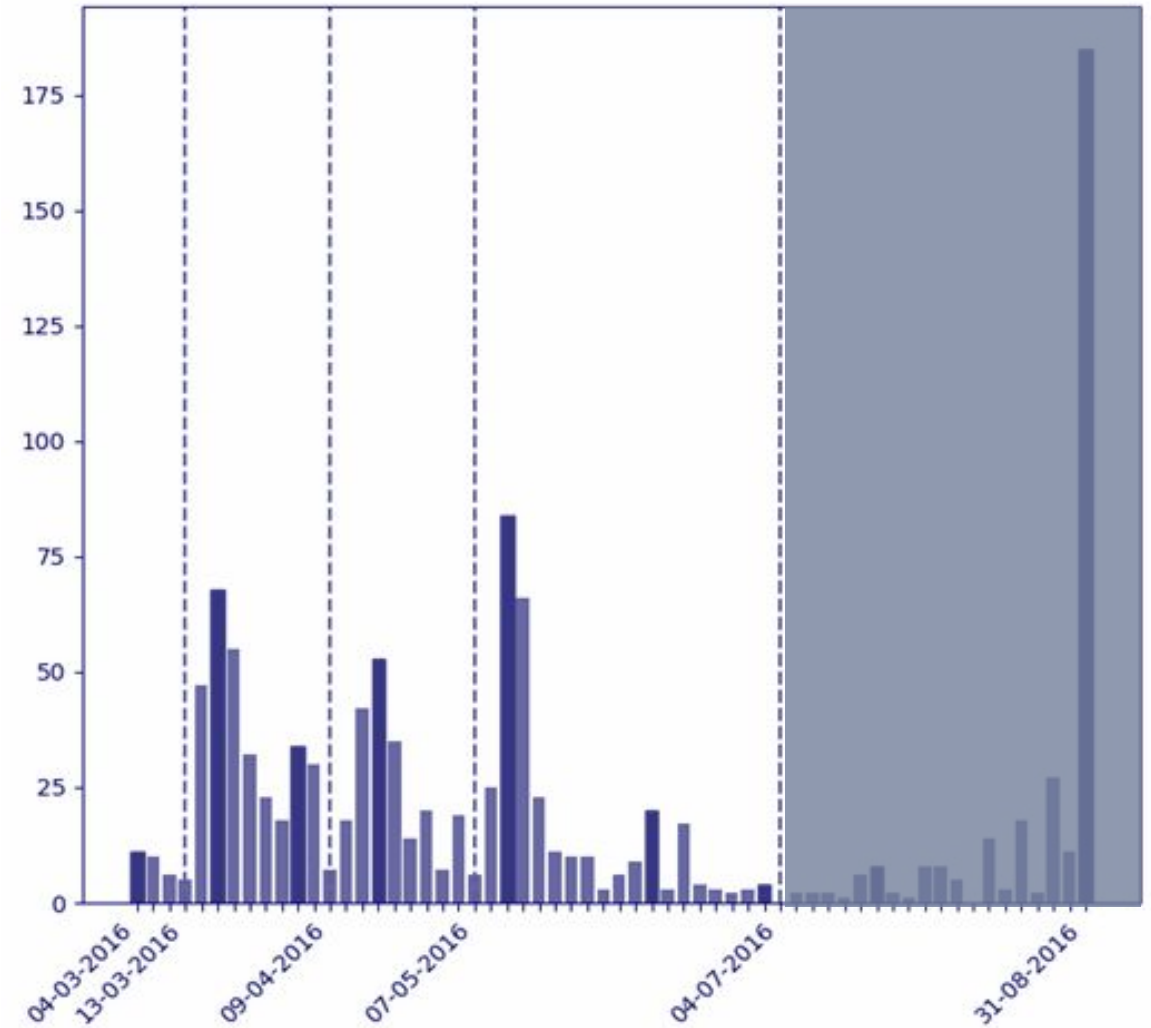


architecture

3.

identify
relevant
time
intervals

peak
detection



architecture



architecture



$$S(kw) = \frac{\prod_{w \in kw} S(w)}{\text{TF}(kw) + (1 + \sum_{w \in ks} S(w))}$$

Considering headlines from each time interval where:

- **S(w)** - term weight
 - Considers weights from every term in the headline
- **TF(kw)** - Headline frequency
 - How many times this headline occurs
- **S(kw)** - Final headline score

More information

Campos, R., & Mangaravite, V., & Pasquali, A., & Jorge, A., & Nunes, C., & Jatowt, A. (2018). A Text Feature Based Automatic Keyword Extraction Method for Single Documents. Best Short Paper of ECIR'18 (<http://bit.ly/YakeDemoECIR2018>)

architecture

5.

user
interface

timeline

dez 2012

nov 2013

fev 2015

- Relevant headlines from **27, April 2014** until **20, February 2015**

Jornalista e fotógrafo do The Times feridos na Síria. negative

China e Rússia vetam proposta para julgar no TPI crimes de guerra na Síria. negative

Guerra síria contamina ainda mais o Líbano. negative

Irão liga ajuda contra Estado Islâmico a negociações nucleares.

ONU acusa Síria e Estado Islâmico de crimes de guerra contra a humanidade. negative

Casa Branca confirmou morte de norte-americano jihadista na Síria. negative

Obama prepara mapa para atacar jihadistas na Síria.

ONU quer governo e Estado Islâmico julgados por crimes de guerra. negative

Obama afasta para já intervenção militar na Síria. negative

Guerra civil na Síria fez mais de três milhões de refugiados. negative

Primeiro-ministro britânico não afasta acção contra Estado Islâmico na Síria. negative

Vídeo com refém britânico compara intervenção na Síria à do Vietname.

Síria diz ter sido informada por Washington dos ataques contra o Estado Islâmico.

Presidente da síria fez rara aparição pública em mesquita de damasco.

Oposição Síria condena decapitação de cidadão britânico. negative

Turquia quer mísseis da NATO apontados para a Síria.

Turcos dão luz verde ao Exército Livre da Síria e travam curdos. positive

Criminoso de guerra nazi Brunner morreu provavelmente na Síria. negative

Terminou o ano com mais mortos na guerra civil síria. negative

more detailed information at

A Text Feature Based Automatic Keyword Extraction Method for Single Documents *

Campos, R., & Mangaravite, V., & Pasquali, A., & Jorge, A., & Nunes, C., & Jatowt, A. (2018). 40th European Conference on Information Retrieval (ECIR'18). Grenoble, France. March 26 – 29. (Vol. 10772(2018), pp. 684 - 691).

YAKE! Collection-independent Automatic Keyword Extractor **

Campos, R., & Mangaravite, V., & Pasquali, A., & Jorge, A., & Nunes, C., & Jatowt, A. (2018). 40th European Conference on Information Retrieval (ECIR'18). Grenoble, France. March 26 – 29. (Vol. 10772(2018), pp. 806 - 810).

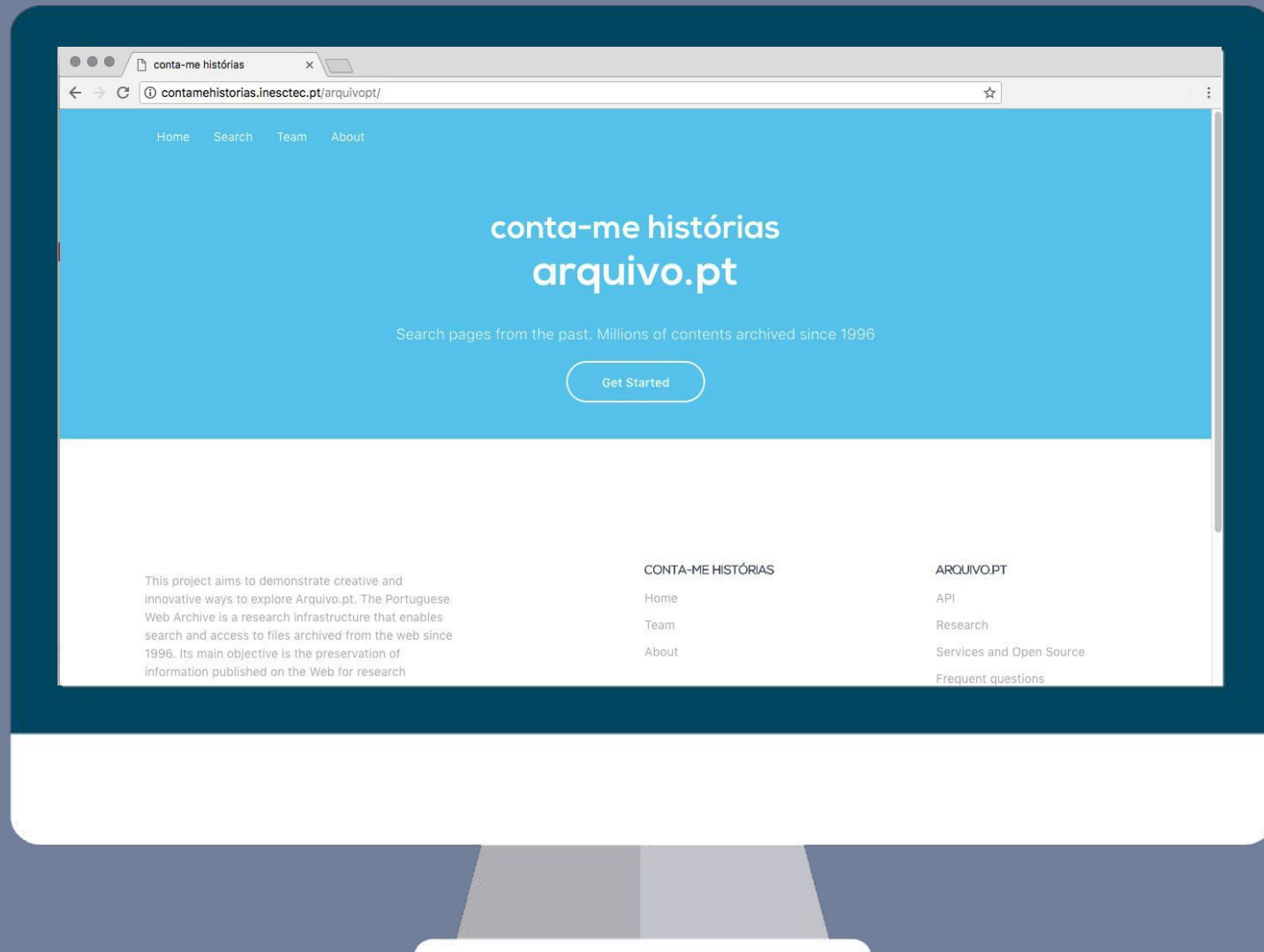


* ECIR 2018 Best Short Paper Award

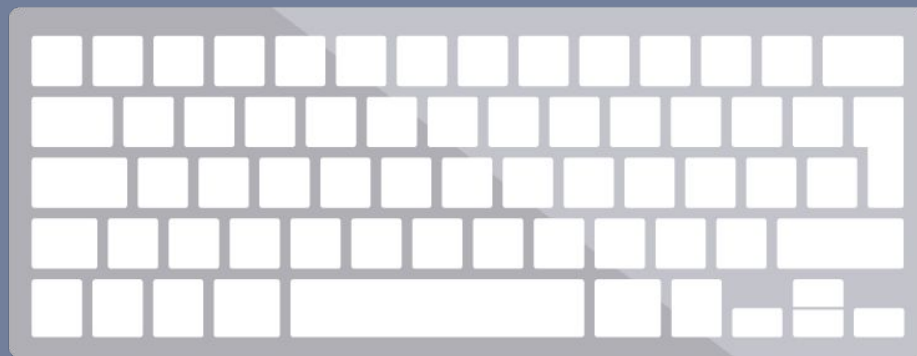
**** Demo**

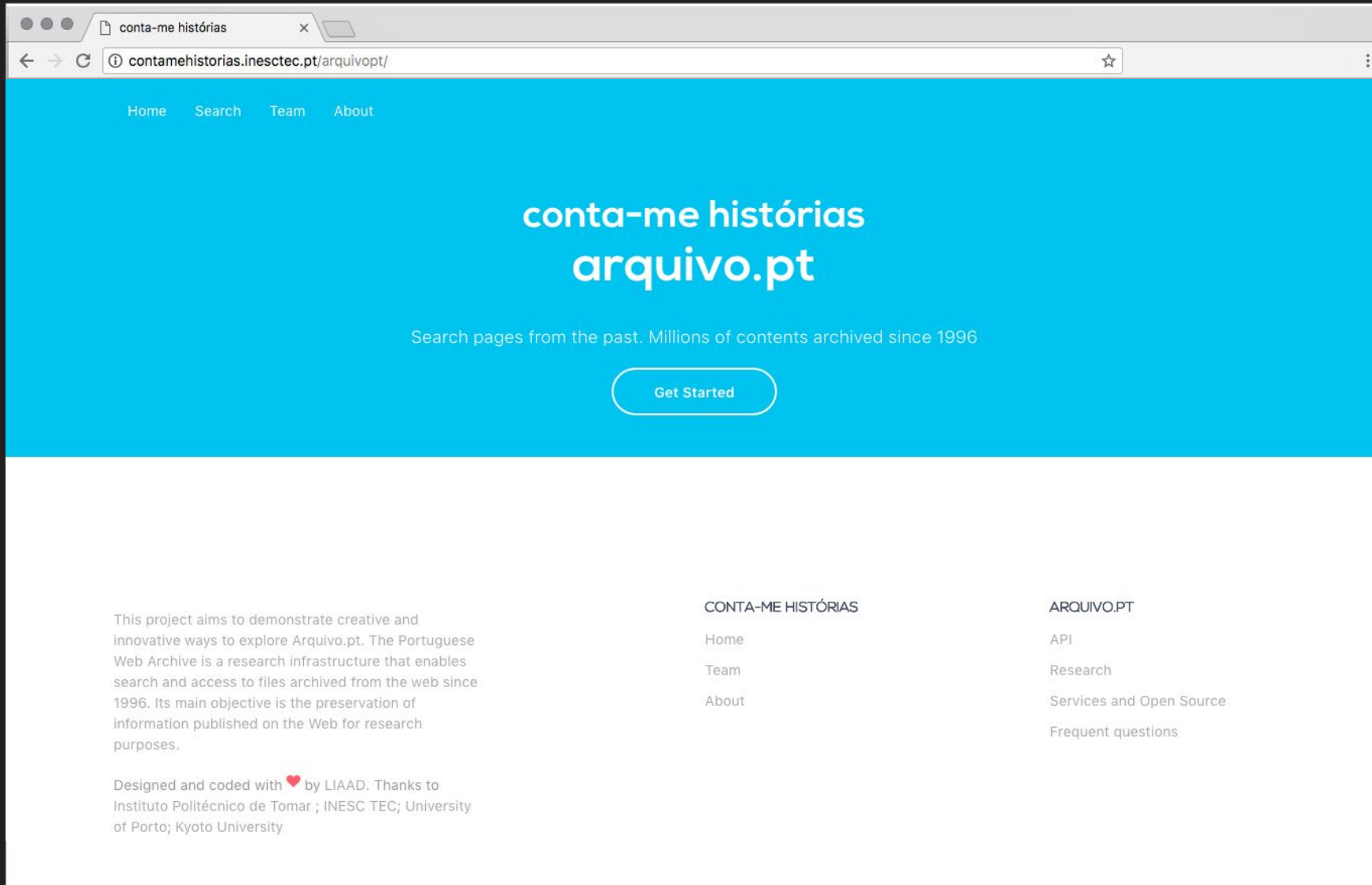
<http://bit.ly/YakeDemoECIR2018>

contamehistorias.pt



<http://contamehistorias.pt>





[Home](#) [Search](#) [Team](#) [About](#)

conta-me histórias arquivo.pt

Search pages from the past. Millions of contents archived since 1996

[Get Started](#)

This project aims to demonstrate creative and innovative ways to explore Arquivo.pt. The Portuguese Web Archive is a research infrastructure that enables search and access to files archived from the web since 1996. Its main objective is the preservation of information published on the Web for research purposes.

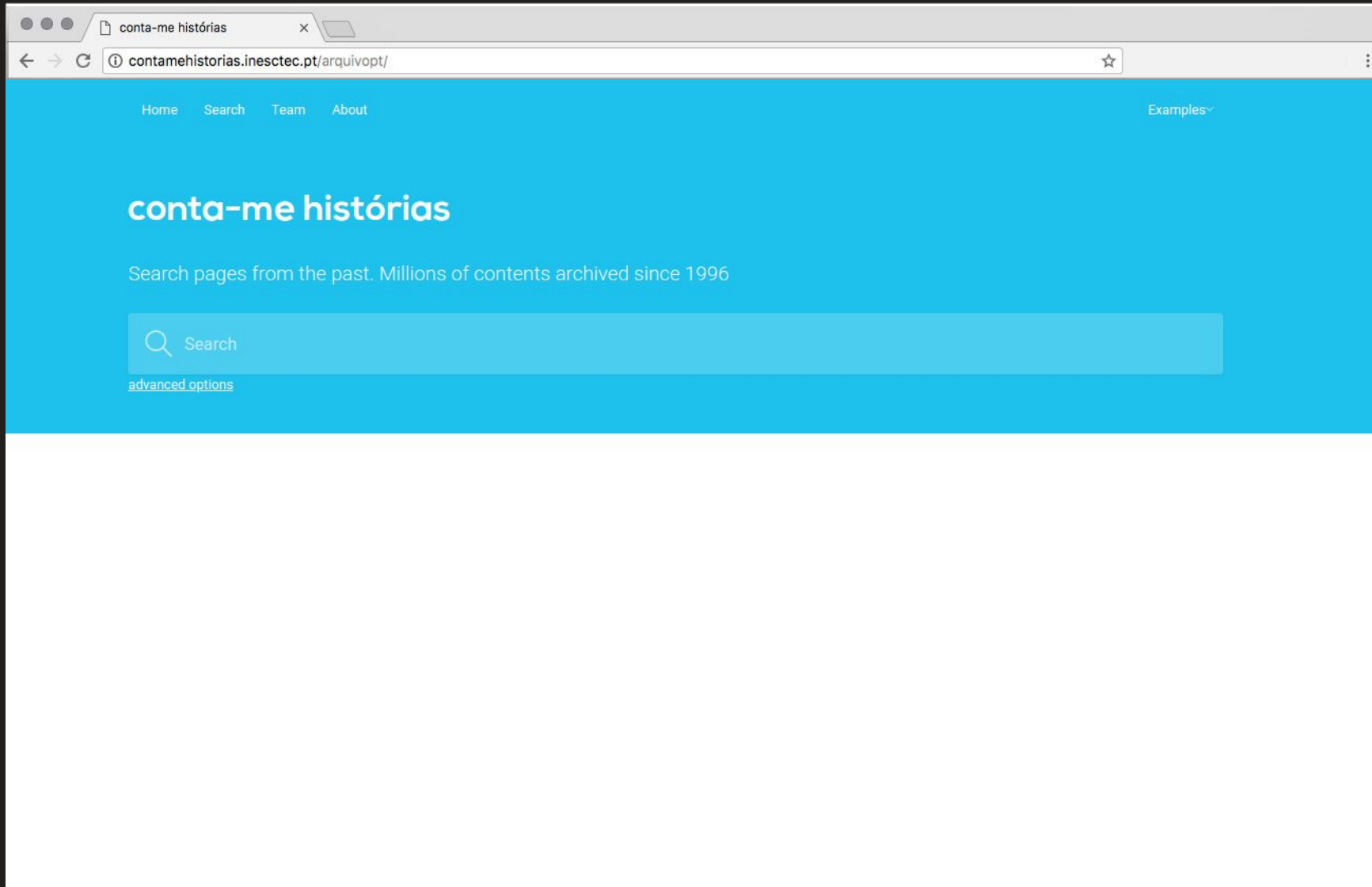
Designed and coded with ❤ by LIAAD. Thanks to Instituto Politécnico de Tomar ; INESC TEC; University of Porto; Kyoto University

CONTA-ME HISTÓRIAS

[Home](#)
[Team](#)
[About](#)

ARQUIVO.PT

[API](#)
[Research](#)
[Services and Open Source](#)
[Frequent questions](#)



conta-me histórias

← → ↻ ⓘ contamehistorias.inesctec.pt/arquivopt/ ☆ ⋮

[Home](#) [Search](#) [Team](#) [About](#) [Examples](#)

conta-me histórias

Search pages from the past. Millions of contents archived since 1996

🔍

troika em portugal

[advanced options](#)

7855
PAGES FOUND

2290
PAGES ANALYZED

17
UNIQUE DOMAINS

[Narrative](#) [Related terms](#)

Narrative Timeline of Troika em Portugal during the last 10 years

Horizontal

Vertical

Previous

Next

conta-me histórias

← → ↻ contamehistorias.inesctec.pt/arquivopt/

Sep 2011 Mar 2012 Mar 2013 Nov 2013 Feb 2015 Nov 2016

- Relevant stories from **02, August 2010** until **27, September 2011**

2011-04-30

Troika vai demolir proteção dos trabalhadores portugueses

positive

PSD apresenta programa eleitoral em Maio

Del Potro vence Estoril Open

Durão Barroso diz que é indispensável que Portugal cumpra objectivos

positive

Desemprego em Espanha baixa em Abril

negative

Portugal deve ir mais além das medidas acordadas com a troika

FIFA confirma que FC Porto é o clube português com mais títulos

FMI e CE preocupados com hipótese de Portugal falhar metas

negative

Manuel Alegre e António Capucho fazem primeira análise do acordo

positive

2011-06-12

Ministros europeus das Finanças aprovam hoje assistência financeira a Portugal

Aprovação de pacote de austeridade é alívio para a Europa e justa para Portugal

positive

Portugal volta hoje ao mercado ensombrado pela Moody

Juros de Portugal dispararam em dia de leilão

António José Seguro

Ferreira do Amaral defende saída digna da moeda única

positive

Portugal recebe cinco ME em compensações europeias

apresenta conclusões de primeira avaliação a Portugal

Mário Soares alerta que a Grécia cair

negative

Eurogrupo expressa confiança na capacidade de Portugal

positive

Sources: Ionline, Tsf, Sabado, Economico Sapo, Dinheirovivo, News Google, Lux lol, Aeiou,

Arquivo.pt: pesquise páginas

arquivo.pt

ARQUIVOPT

2011

Maio

Junho

2

5

7

9

10

12

00:06

16

19

20

21

22

23

24

00:06

www.tsf.pt/PaginalInicial/E..._id=1853012

TSF

RÁDIO NOTÍCIAS

Actualidade

Programas

Multimédia

Blogues

Futebol

Bolsa

Trânsito

Totojogos

Meteorologia

Marés

Farmácias

Agenda

Jogos

PSI20

PSI Geral

IBEX35

CAC 40

Euronext

FTSE

NASDAQ 100

Outros Mercados

pesquisar

E.coli

Christine Lagarde

dívida

Eurogrupo

Calçado

FMI

Grécia

sindicatos

Noticiário das 00h00

Fórum TSF

Podcast

RSS

FM

MOBILE

SMS

Fale connosco

login | registo

Economia

Administração Pública

Orçamento

Bolsa

Empresas

Patrocínio

Ferramentas

Enviar

Partilhar

Imprimir

Aumentar

Diminuir

Comentar este Artigo

Estatísticas

515 Visualizações

0 Comentários

0 Envios

Ministros europeus das Finanças aprovam hoje assistência financeira a Portugal

16 MAI 11 às 06:29

Os ministros das Finanças da Zona Euro e da União Europeia deverão aprovar sem dificuldade, esta segunda-feira em Bruxelas, o programa de assistência financeira a Portugal, depois de ter sido afastada a ameaça da Finlândia se opor à decisão.

O acordo político deverá ser alcançado numa reunião, esta tarde às 17h00 (16h00 de Lisboa), dos responsáveis da Zona Euro alargada aos ministros da UE (os 17 anteriores mais dez que não pertencem à Zona Euro).

A assistência financeira a Portugal será repartida em partes iguais de 26 mil milhões de euros pelo Fundo Europeu de Estabilização Financeira (FEFF).

Últimas

+ Vistas

+ Comentadas

+ Pesquisadas

Projecto Novos Povoadores é alternativa à vida na cidade

Ontem às 23:44

Governador do Banco de Israel é o novo candidato ao FMI

Ontem às 23:16

Explosões no Paquistão causam 25 mortos e 100 feridos

Ontem às 22:50

Cáritas pede criação de rede de protecção social

Ontem às 22:39

Blue Note é a Melhor Editora de Jazz do Mundo

Ontem às 22:24

conta-me histórias

contamehistorias.inesctec.pt/arquivopt/

Sep 2011Mar 2012Mar 2013Nov 2013Feb 2015Nov 2016

- Relevant stories from **01, July 2012** until **24, March 2013**

Avaliação positiva da Troika contribui para melhorar imagem do país

positive

Sentiment automatically detected based on the words that are part of the sentence

Seleccção nacional de regresso a Portugal

FMI diz que vai discutir mais austeridade na próxima avaliação da troika

positive

Portugal devia mostrar relatório sobre pobreza à troika

negative

Troika regressa a Portugal em semana dominada por Miguel Relvas

negative

Estela Barbot acredita que troika terá bom senso na próxima avaliação

positive

Deputado do CDS pede que troika corte nos juros

Berlim e EUA saudam esforços de Portugal para cumprir programa

Portugal merece ser recompensado pela troika

Passos Coelho não garante que IRS não vá aumentar

Troika obriga Portugal a pagar juros de dinheiro que não utiliza

Juros da dívida sobem em Portugal e Espanha

Juros da dívida aliviam em Portugal antes de leilão

Angela Merkel convencida de que Portugal vai cumprir memorando

positive

Gaspar cita Argentina para dizer que Portugal tem de honrar compromissos

positive

Portugal já merecia uma atenção especial da troika

positive

Portugal desaconselhado a pedir condições iguais às da Grécia

Portugal Ways Lounge

Portugal conquista três prémios no turismo mundial

positive

Presidente da AICEP defende mais Economia e menos Finanças

Sources: Ionline, Tsf, Sabado, Economico Sapo, Dinheirovivo, News Google, Lux lol, Aeiou,

conta-me histórias x

← → ↻ ⓘ contamehistorias.inesctec.pt/arquivopt/ ☆ ⋮

Sep 2011 Mar 2012 Mar 2013 Nov 2013 Feb 2015 Nov 2016

- Relevant stories from **06, February 2015** until **16, November 2016**

2015-02-28 Portugal reembolsará seis mil milhões de euros ao FMI em Março

Paulo Portas diz ter havido uma troika de restrições e uma troika de bem-fazer

Portugal não vai recuperar nem metade dos empregos destruídos pela troika positive

Troika culpa Governo português por falhanço do programa negative

Passos e Cavaco são a troika em Portugal

Parlamento Europeu vai investigar erros da troika em Portugal positive

Troika regressa a Portugal na pr³xima semana

Troika regressa a Lisboa para segunda avaliação pós-resgate

Paulo Portas numa feira em Bruxelas lembra que Portugal não é a Grécia

chega hoje a Lisboa para debater TSU e Orçamento do Estado

Troika volta a Portugal e não é para matar saudades

Costa diz que visita da troika a Portugal

Banda desenhada do vinho do Porto na Garrafeira Nacional

The Guardian e Lonely Planet escolhem as melhores dicas sobre Lisboa

Germano Silva vai ser Doutor Honoris Causa pela Universidade do Porto positive

Atletas gregos sem dinheiro lutam para competir nos Jogos do Rio de Janeiro positive

Banco de Portugal e CMVM

Troika em Portugal para analisar contas p^oblicas

Presidente da Águas de Portugal demite-se em ruptura com o Governo

Carlos Costa diz que CMVM devia ser informada pelo BES

Sources: Ionline, Tsf, Sabado, Economico Sapo, Dinheirovivo, News Google, Lux lol, Aeiou,



01/01/1996

31/12/2017

Pesquisa avançada

Lista de versões

1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
13 Out		12 Dez	25 Jan	11 Mai	18 Abr	30 Mar	8 Fev	2 Jan	4 Jan	1 Jan	2 Fev	1 Jan	20 Mai	23 Mar	1 Jan	1 Jan	10 Jan	27 Abr	1 Jan	1 Jan	Disponíveis após 1 ano.
13 Out		12 Dez	25 Jan	11 Mai	3 Mai	25 Mai	18 Fev	6 Fev	15 Jan	4 Jan	16 Mar	1 Jan	24 Mai	1 Abr	1 Jan	1 Jan	10 Jan	27 Abr	1 Jan	1 Jan	
			8 Fev	11 Mai	6 Mai	28 Mai	22 Mar	11 Abr	17 Jan	4 Jan	16 Mar	5 Jan	24 Jun	5 Abr	2 Jan	2 Jan	11 Jan	27 Abr	1 Jan	1 Jan	
			8 Fev	11 Mai	6 Mai	1 Jun	26 Mar	8 Mai	18 Jan	4 Jan	6 Jun	11 Jan	24 Jun	8 Abr	2 Jan	2 Jan	11 Jan	27 Abr	2 Jan	1 Jan	
			21 Fev	11 Mai	16 Mai	4 Jun	4 Abr	20 Mai	20 Jan	6 Jan	7 Jun	15 Jan	25 Set	9 Abr	3 Jan	3 Jan	12 Jan	28 Abr	2 Jan	2 Jan	
			21 Fev	11 Mai	22 Mai	6 Jun	11 Abr	23 Mai	22 Jan	7 Jan	7 Jun	18 Jan	26 Set	12 Abr	3 Jan	3 Jan	12 Jan	28 Abr	2 Jan	2 Jan	
			21 Abr	11 Mai	8 Jun	2 Ago	24 Abr	26 Mai	24 Jan	8 Jan	7 Jun	20 Jan	17 Dez	13 Abr	4 Jan	4 Jan	15 Jan	28 Abr	3 Jan	2 Jan	
			21 Abr	11 Mai	27 Jun	7 Ago	25 Mai	6 Jun	27 Jan	11 Jan	7 Jun	23 Jan	18 Dez	16 Abr	4 Jan	4 Jan	15 Jan	28 Abr	3 Jan	2 Jan	
			27 Abr	11 Mai	27 Jun	12 Ago	17 Jul	7 Jun	28 Jan	13 Jan	9 Jun	26 Jan		19 Abr	5 Jan	5 Jan	17 Jan	29 Abr	3 Jan	3 Jan	
			27 Abr	11 Mai	28 Jun	12 Ago	30 Set	10 Jun	28 Jan	14 Jan	14 Jun	1 Fev		22 Abr	5 Jan	5 Jan	17 Jan	29 Abr	4 Jan	3 Jan	
			12 Out	11 Mai	28 Jun	23 Set	10 Out	10 Jun	29 Jan	14 Jan	14 Jun	3 Fev		30 Abr	6 Jan	6 Jan	18 Jan	29 Abr	4 Jan	3 Jan	
			12 Out	20 Mai	1 Jul	28 Set	21 Out	11 Jun	30 Jan	14 Jan	6 Ago	9 Fev		2 Mai	6 Jan	6 Jan	18 Jan	30 Abr	4 Jan	3 Jan	
			11 Nov	20 Mai	9 Jul	24 Out	24 Nov	12 Jun	31 Jan	15 Jan	27 Set	13 Fev		3 Mai	7 Jan	7 Jan	19 Jan	30 Abr	5 Jan	4 Jan	

conta-me histórias

contamehistorias.inesctec.pt/arquivopt/

☆

Narrative

Related terms

Narrative Timeline of Troika Em Portugal during the last 10 years

Horizontal

Vertical

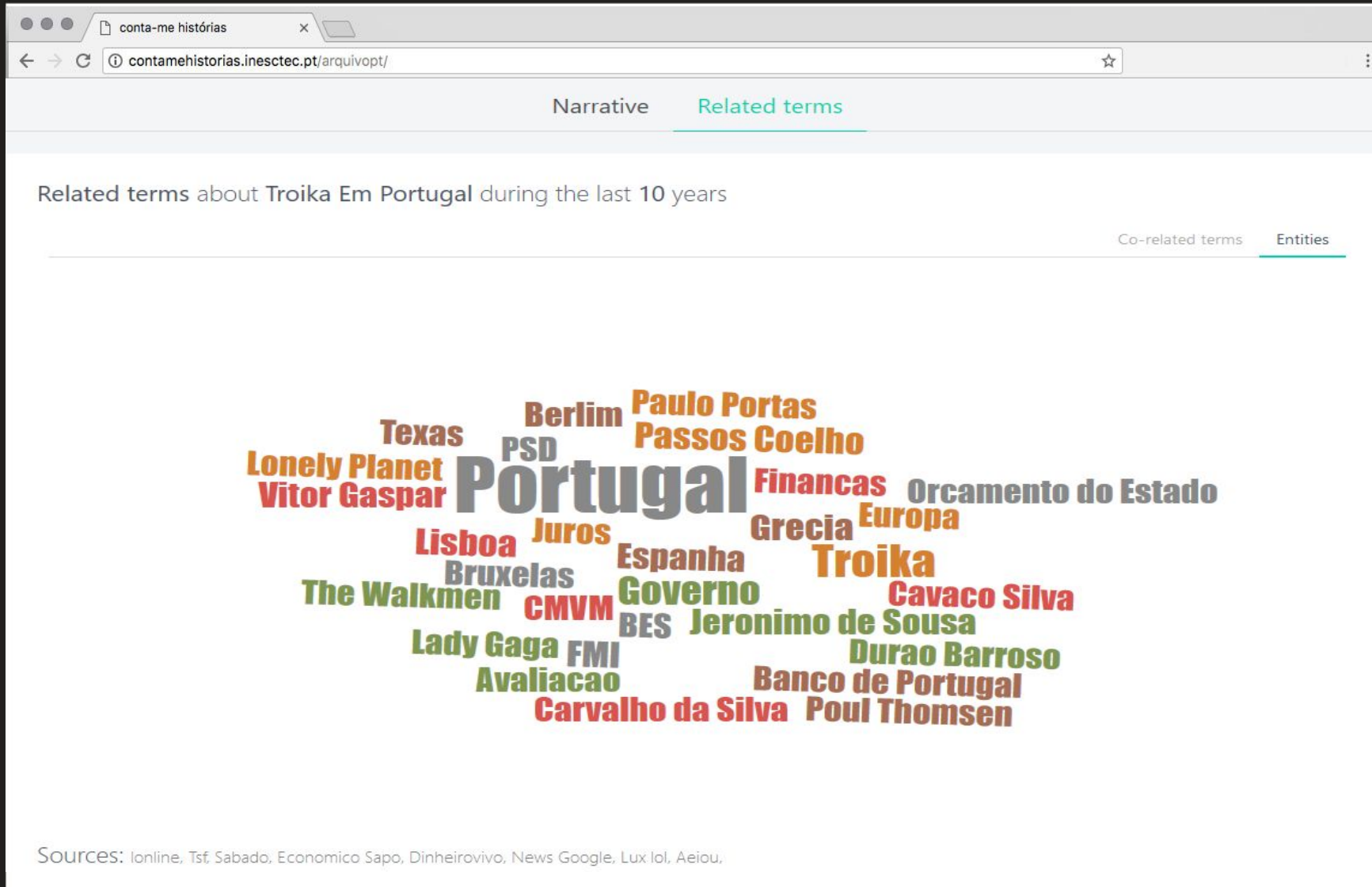
16, November 2016

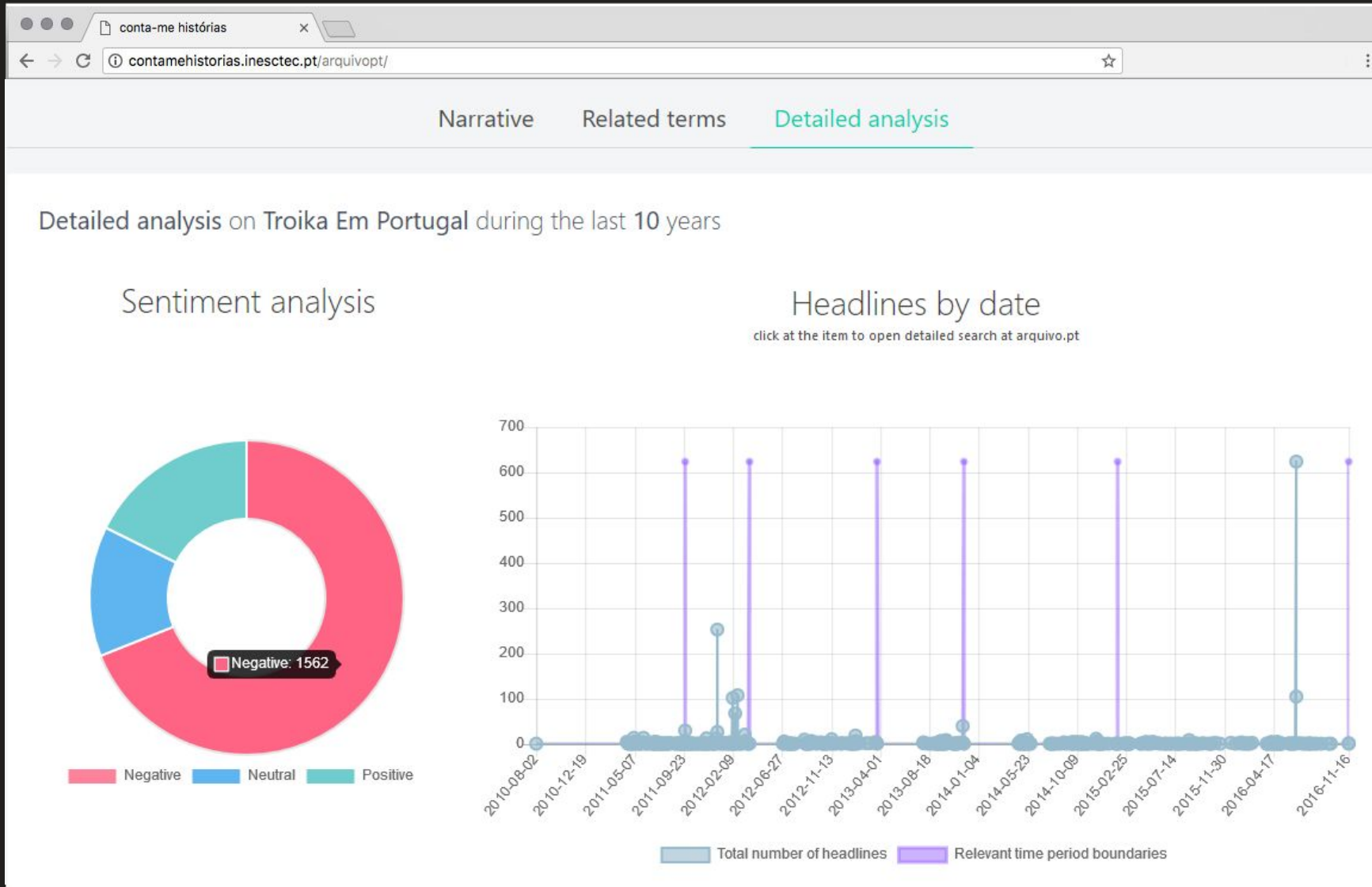
- Portugal reembolsará seis mil milhões de euros ao FMI em Março
- Paulo Portas diz ter havido uma troika de restrições e uma troika de bem-fazer
- Portugal não vai recuperar nem metade dos empregos destruídos pela troika
- Troika culpa Governo português por falhanço do programa
- Passos e Cavaco são a troika em Portugal
- Parlamento Europeu vai investigar erros da troika em Portugal

2 years ago

01, February 2015

- Troika diz que Portugal precisa
- Última avaliação da troika arranca hoje em Lisboa
- Jerónimo de Sousa diz que saída da troika de






conta-me histórias

Arquivo.pt: pesquise páginas

arquivo.pt



Troika em Portugal

Pesquisar

[Pesquisa avançada](#)

entre: 27/12/2011 e: 27/12/2011

Resultados 1 a 10 de 14.911

[Troika: Portugal «tem feito bons progressos em várias frente...](#)

27 Dezembro, 2011 - [Listar versões](#)

Troika: Portugal «tem feito bons progressos **em** várias frentes» - Comissão Europeia | União Europeia ... : **Portugal** «tem feito bons progressos **em** várias frentes» - Comissão Europeia A Comissão Europeia considerou hoje que **Portugal** «tem feito bons progressos **em** várias frentes» no que refere à implementação do ...

http://diariodigital.sapo.pt/news.asp?section_id=57&id_news=549293

[Troika: Portugal não deve precisar de mais dinheiro que o in...](#)

27 Dezembro, 2011 - [Listar versões](#)

Troika: Portugal não deve precisar de mais dinheiro que o inicialmente acordado - Bruxelas | União Europeia | Diário Digital Director: Pedro Curvelo Pesquise no DD **em** Todas as secções.. Sociedade ... **Troika: Portugal** não deve precisar de mais dinheiro que o inicialmente acordado - Bruxelas A Comissão ...

http://diariodigital.sapo.pt/news.asp?section_id=57&id_news=549315

[Portugal na "lupa" da troika](#)

27 Dezembro, 2011 - [Listar versões](#)

Portugal na "lupa" da **troika** Skip to content Skip to main navigation Skip to 1st column Skip to 2nd ... Economia

Portugal na "lupa" da **troika** **Portugal** na "lupa" da **troika** Terça, 08 Novembro 2011 00:00 Redacção

TRANSLATE english french spanish swahili **Portugal** na "lupa" da **troika** **Portugal** na "lupa" ...

<http://opais.sapo.mz/index.php/economia/...17575-portugal-na-lupa-da-troika.html>

[Sol - Dossier - troika](#)

27 Dezembro, 2011 - [Listar versões](#)

Economia **Portugal** dever♦ poupar 5,2 mil mil♦es **em** juros durante ajuda externa A Comiss♦o Europeia ... ♦

♦ comentar Economia **Troika: Portugal** n♦o deve precisar de mais dinheiro A Comiss♦o Europeia ... **'troika'** no programa de ajustamento... ♦ ♦ comentar Economia Rendimento dispon♦vel cai 6% **em** 2012 O ...

conta-me histórias

conta-mehistorias.inspector.pt/arquivo.pt

Narrative

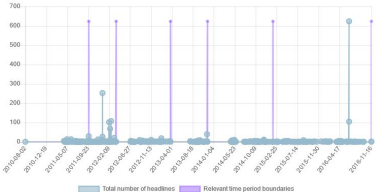
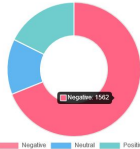
Related terms

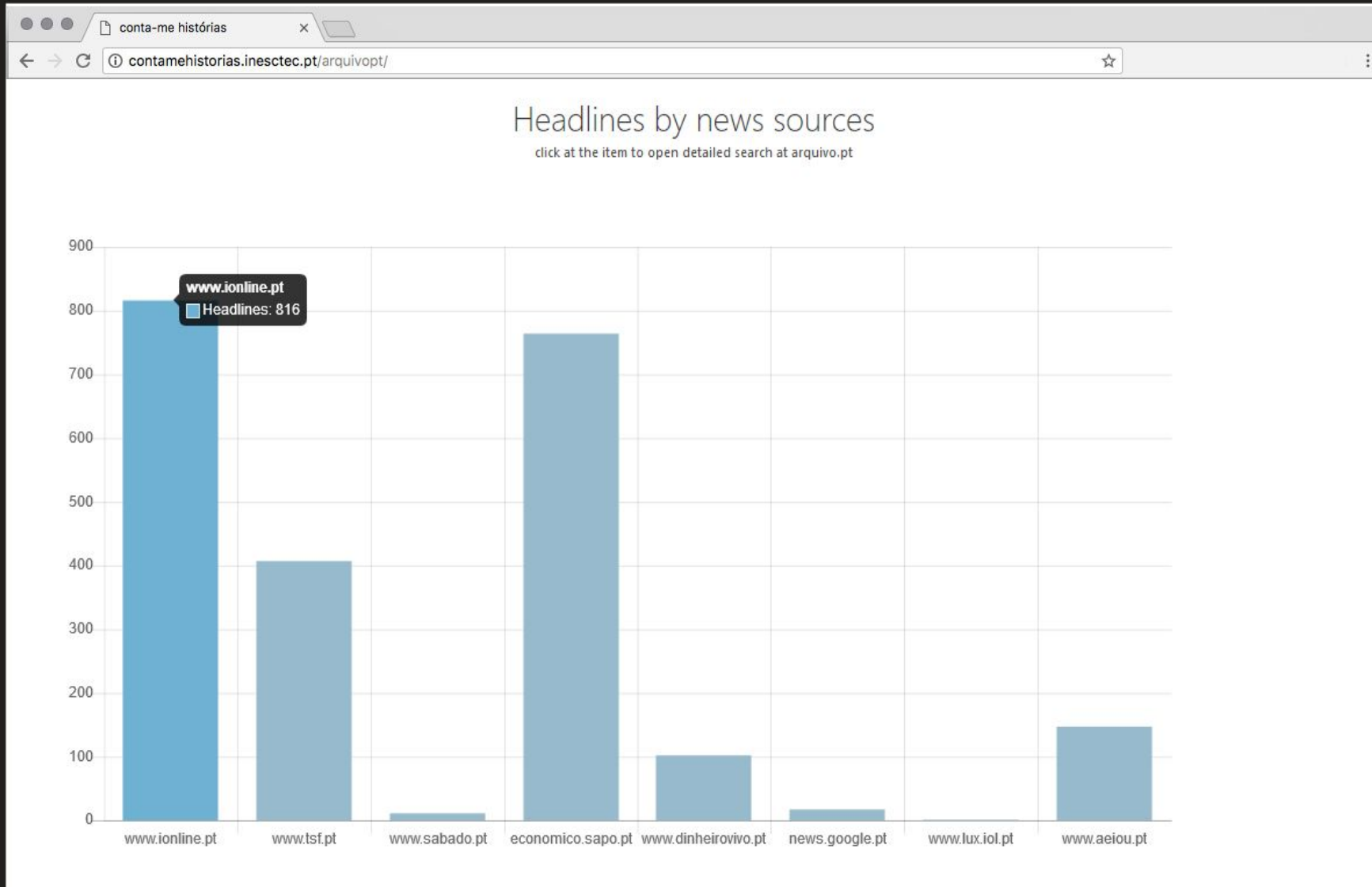
Detailed analysis

Detailed analysis on Troika Em Portugal during the last 10 years

Sentiment analysis

Headlines by date






conta-me histórias

Arquivo.pt: pesquise páginas

arquivo.pt



Troika em Portugal site:www.ionline.pt type:html

Pesquisar

[Pesquisa avançada](#)

entre: 02/08/2010 e: 16/11/2016

Resultados 1 a 10 de 346.782

Desejava ver apenas o [histórico de www.ionline.pt?](#)

[Troika: Portugal à espera dos próximos episódios](#)
24 Abril, 2011 - [Listar versões](#)
Troika: Portugal à espera dos próximos episódios Fechar Imprimir Troika: Portugal à espera dos próximos episódios Publicado em 22 de Abril de 2011 A troika começou as reuniões com os partidos e ... Impostas pela troika aos portugueses deverão ser conhecidas na próxima semana. Ainda uma viagem à ...
<http://www.ionline.pt/interior/index.php?p=news-print&idNota=118903>

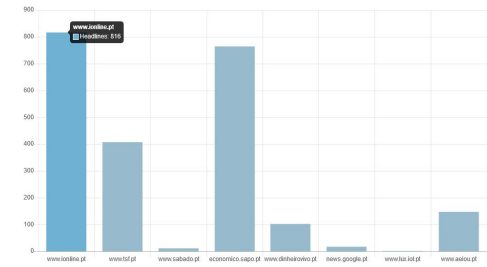
[Costa. A troika em Portugal agora chama-se Portas, Passos e ...](#)
24 Novembro, 2015 - [Listar versões](#)
Costa. A troika em Portugal agora chama-se Portas, Passos e Cavaco 24/11/15 subscrever newsletter ... Registrar
Pesquisar Costa. A troika em Portugal agora chama-se Portas, Passos e Cavaco Costa. A troika em Portugal agora chama-se Portas, Passos e Cavaco ANDRÉ KOSTERS, Lusa António Costa 23/04 ...
<http://www.ionline.pt/388076>

[Troika. Portugal vai ao terceiro exame da cadeira Maior Ajus...](#)
12 Fevereiro, 2012 - [Listar versões](#)
Troika. Portugal vai ao terceiro exame da cadeira Maior Ajustamento Económico de sempre | iOnline ... Mercados Boa Vida iModa iTestdrive iGastronomia Desporto ITV iFotogaleria Opinião iSabe Troika. Portugal ... Portugal, os bancos podem cumprir a meta já em 2013. Mas há um preço: o esmagamento do crédito empresarial ...
<http://www.ionline.pt/dinheiro/troika-po...ra-maior-ajustamento-economico-sempre>

[Troika: Portugal não deve precisar de mais dinheiro que o in...](#)
22 Dezembro, 2011 - [Listar versões](#)
Troika: Portugal não deve precisar de mais dinheiro que o inicialmente acordado - Bruxelas ... iTestdrive

Headlines by news sources

click at the item to open detailed search at arquivo.pt



News Source	Headlines Count (approx.)
www.ionline.pt	800
www.tsf.pt	400
www.sabado.pt	20
economico.sapo.pt	750
www.dinheiro.pt	100
news.google.pt	20
www.lusa.pt	10
www.aula.pt	150

conta-me histórias

← → ↻ ⓘ contamehistorias.inesctec.pt/arquivopt/ ☆ ⋮

Narrative Timeline of Donald Trump during the last 10 years

Horizontal

Vertical

Previous

Next

Jan 2018

Feb 2018

Mar 2018

Apr 2018

Apr 2018

May 2018

May 2018

- Relevant stories from 06, February 2018 until 20, March 2018

Trump Plan to Move Drug Prevention Program to Justice Dept positive

Trump Expected to Release Democratic Memo

Trump 's Infrastructure Plan May Ignore Climate Change

American ISIS Suspect Said He Wanted to Report From Syria

Russia Wanted Trump to Win positive

Trump Edges Toward a Conventional Republican Approach

Trump to Discuss School Shootings With Survivors

Trump Condemns Florida Officer Who Did n't Act negative

Trump Blames Video Games for Mass Killings negative

Trump Embraces a Trade War negative

Trump Advisers Fervently Defend Tariffs

South Korean Official on Trump

Trade War Prospect Shakes Part of Trump Base

Trump to Push Ahead on Gun Training for School Employees

Trump Plans to View Wall Prototypes

Trump Finds Unlikely Culprit in School Shootings

Trump Says He Made Up Trade Numbers in Talk with Trudeau

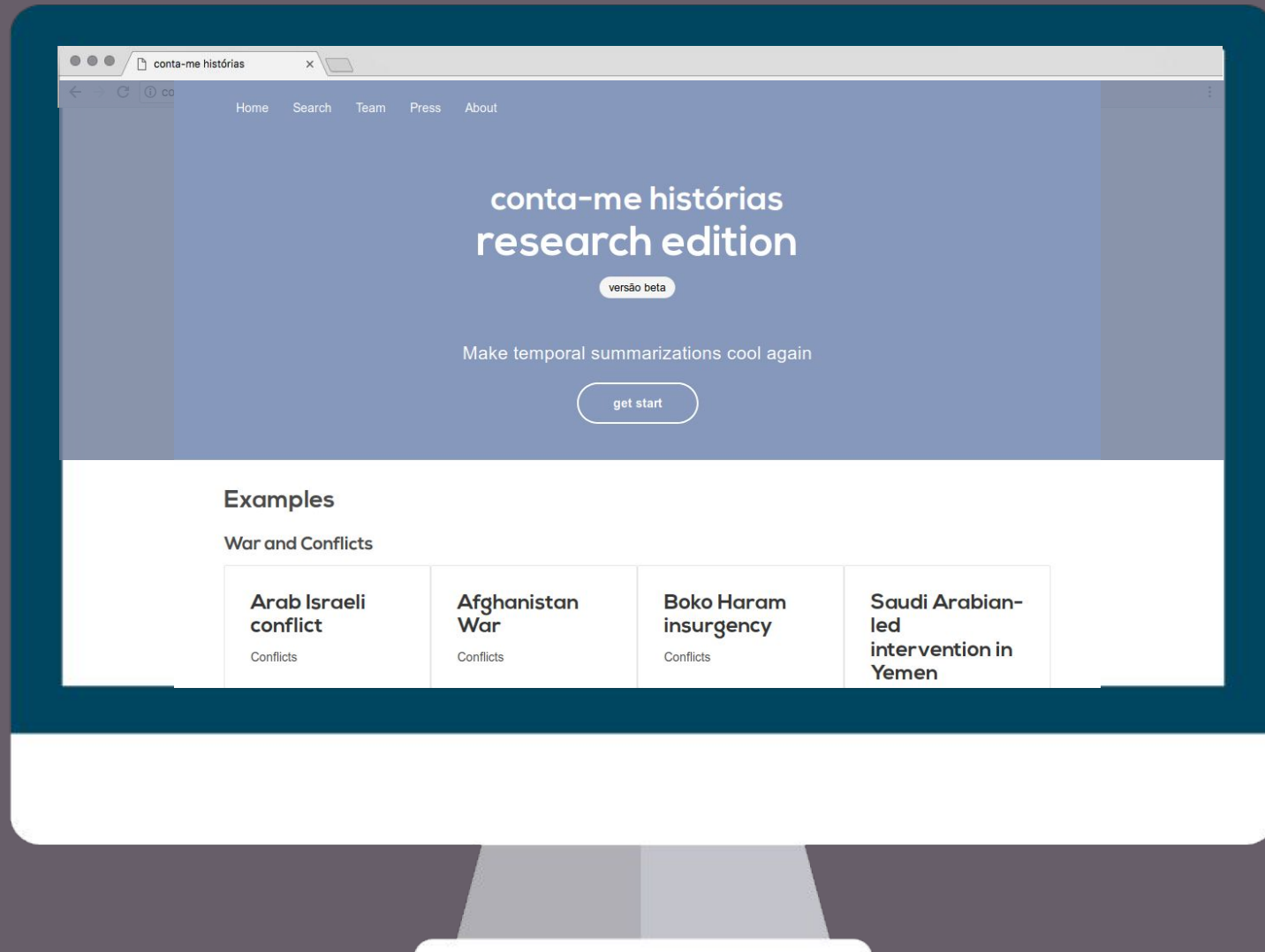
Assaults Increased When Cities Hosted Trump Rallies

Trump to Hire Lawyer Who Has Pushed Theory That Justice Dept positive

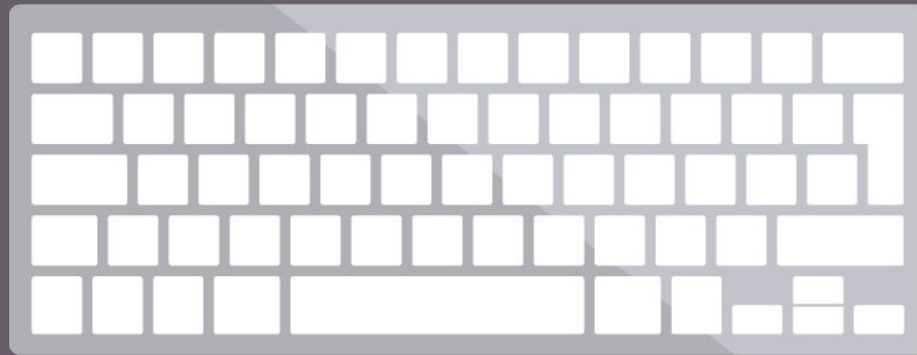
Melania Trump Wants to End Online Bullying negative

Sources: New York Times,

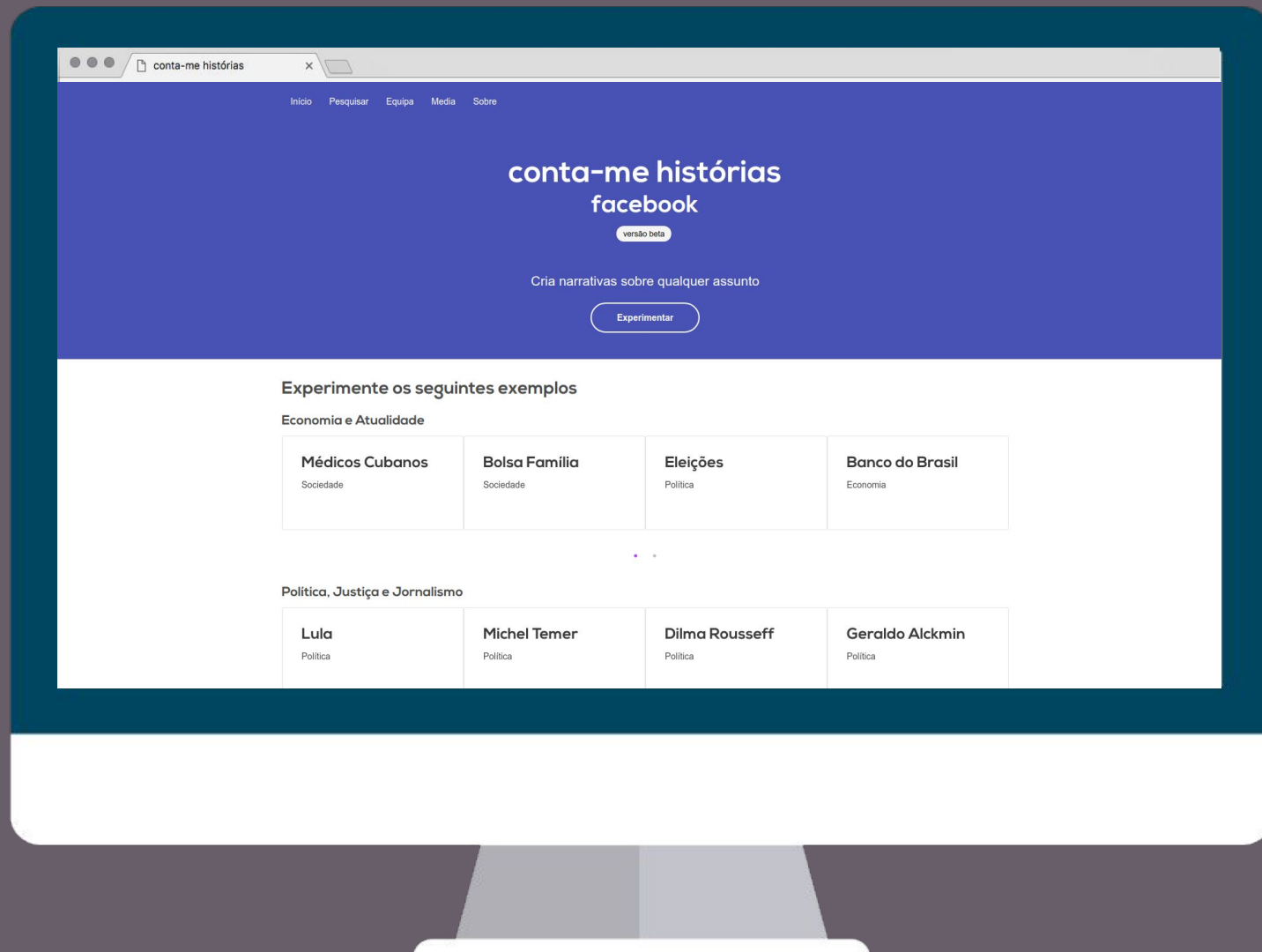
research edition with different datasets



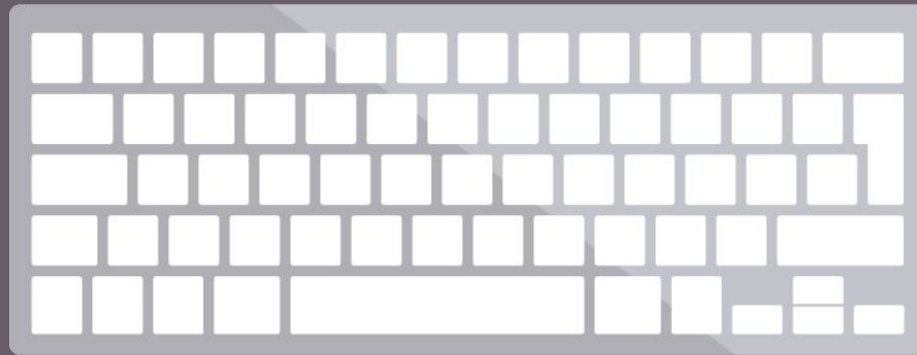
<http://nlplab.inesctec.pt/contamehistorias-labs>



facebook posts edition



<http://nlplab.inesctec.pt/contamehistorias-facebook>



Conta-me Histórias open source

open source python package

Search or jump to... Pull requests Issues Marketplace Explore

Your repository details have been saved.

LIAAD / TemporalSummarizationFramework

Unwatch 4 Star 1 Fork 0

<> Code Issues 1 Pull requests 0 Projects 0 Wiki Insights Settings

Conta-me Histórias, temporal summarization engine <http://contamehistorias.pt> Edit

Manage topics

45 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

arianpasquali Update LICENSE.txt Latest commit a08fc6a 3 hours ago

contamehistorias	refactoring e levenshtein	a day ago
LICENSE.txt	Update LICENSE.txt	3 hours ago
MANIFEST.in	refactoring e levenshtein	a day ago
README.md	Update README.md	7 days ago
requirements.txt	refactoring e levenshtein	a day ago
setup.cfg	terminal script and pip install	17 days ago
setup.py	refactoring and new examples	15 days ago
test_conteme.py	terminal script and pip install	17 days ago

<http://github.com/LIAAD/TemporalSummarizationFramework>

Installing

requires Python 3



```
> pip install git+https://github.com/LIAAD/TemporalSummarizationFramework.git
```


Using terminal interface

Accessing client help



```
> contamehistorias --help
```

```
Usage: contamehistorias [OPTIONS]
```

```
Console script for contamehistorias.
```

```
Options:
```

<code>--query TEXT</code>	Perform news retrieval with informed query
<code>--language TEXT</code>	Expected language in headlines
<code>--start_date DATE</code>	Perform news retrieval since this date
<code>--end_date DATE</code>	Perform news retrieval until this date
<code>--domains TEXT</code>	Comma separated list of domains (www.publico.pt,www.dn.pt)
<code>--verbose</code>	
<code>--help</code>	Show this message and exit.

Code with Arquivo.pt as data source



```
from contamehistorias.datasources.webarchive import ArquivoPT
from contamehistorias.engine import TemporalSummarizationEngine
from datetime import datetime

domains = [ 'http://publico.pt/', 'http://www.dn.pt/', 'http://www.rtp.pt/', 'http://www.iol.pt/' ]

params = { 'domains':domains, 'from':datetime(year=2016, month=3, day=1),
          'to':datetime(year=2018, month=1, day=10) }

query = 'Dilma Rousseff'
language = "pt"

#instantiate ArquivoPT search engine
apt = ArquivoPT()
search_result = apt.getResult(query=query, **params)
#instantiate temporal summarization class
engine = TemporalSummarizationEngine()
intervals = cont.build_intervals(search_result, language)
#print results
engine.pprint(intervals)
```

How to extend it

You can extend it to use your own collection as data source.


Extend BaseDataSource class and implement method **getResult** to return a list of object **ResultHeadLine**.

For each document you need a **title**, a **timestamp**, **source** and **url**.



```
class BaseDataSource(object):  
  
    def __init__(self, name):  
        self.name = name  
  
    def getResult(self, query, **kwargs):  
        raise NotImplementedError('getResult on ' + self.name + ' not implemented yet!')
```

Extension example with custom dataset *



```
from contamehistorias.engine import TemporalSummarizationEngine
from contamehistorias.datasources.signal import SignalNewsIRDataset

#collect dataset
query = "Barack Obama"
search_result = SignalNewsIRDataset().getResult(query=query)

#generate timeline
engine = TemporalSummarizationEngine()
summary_result = engine.build_intervals(search_result, "en")

#print results
engine.pprint(summary_result, verbose=True)
```

* Signal's NewsIR'16 dataset : <http://research.signalmedia.co/newsir16/signal-dataset.html>

Output



```
> contamehistorias_signal --query "Barack Obama"
```

```
Conta-me Historias Temporal Summarization. Signal NewsIR 16 dataset example (more info  
http://research.signalmedia.co/newsir16/signal-dataset.html  
--query "Barack Obama"
```

```
Performing search. This may take a while.  
Computing temporal summarization ...
```

```
Timeline
```

```
2015-07-25 08:44:54 until 2015-09-07 06:53:14
```

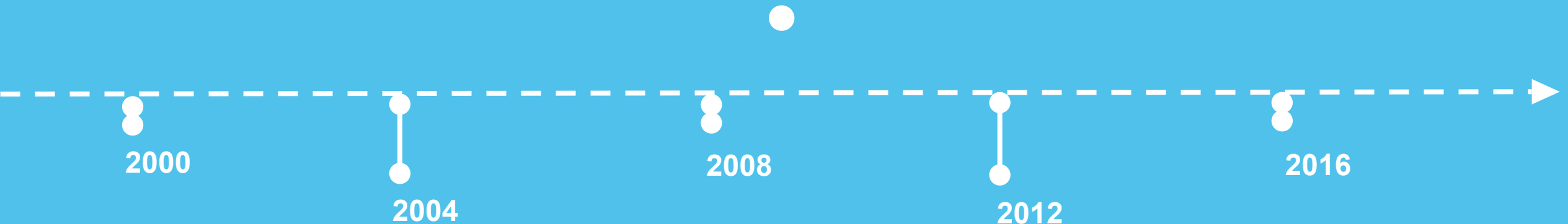
```
  obama says world must reach climate deal in paris  
  obama enthusiastic for los angeles summer olympics bid  
  democrats give obama votes to uphold iran nuke deal  
  obama scores policy win in securing votes for iran nuclear deal
```

```
2015-09-14 04:11:35 until 2015-09-21 22:27:54
```

```
  president obama honors stephen kin with the national medal of the arts  
  struggling gop presidential candidate scott walker blames obama for police deaths  
  obama unveils new college financial aid application process  
  obama will greet chinese president with handshake  
  barack obama supports ahmed mohamed
```

next steps

- In terms of data sources support different web archives
- Different timelines for different kinds of sources (left-wing, right-wing, tabloids, etc)
- Extended formal evaluation





thank you

Arian Pasquali
arianpasquali@gmail.com
INESCTEC - LIAAD
Universidade do Porto

<http://contamehistorias.pt>
admin@contamehistorias.pt