

Online annotation platform for web archives

Pedro Gomes, Arquivo.pt, pedro.gomes@fccn.pt



Arquivo.pt preserves national and international historical web content

GALILEO Versão 2.0

Janeiro 1993

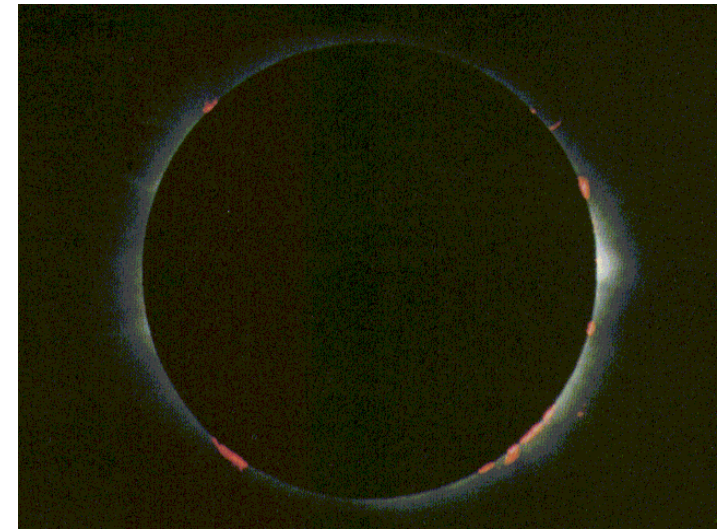
Manual de utilização e sugestões de exploração	<i>AUTORES</i> Programação: João Veloso Manual: Elisa Prata Pina e M. Augusta Patricio Orientação: Carlos Fiolhais
--	---

Departamento de Física da Universidade de Coimbra

ÍNDICE

1. Apresentação sumária do programa
2. Características do equipamento
3. Configuração mínima do computador
4. Ficheiros do programa

nautilus.fis.uc.pt- 1993
(oldest page)



spacelink.nasa.gov – 1992
(oldest image)

How to **search** and **access**
the preserved information?



Search **texts** from the past in any language

The screenshot displays the ARQUIVO.PT search interface. At the top, there is a dark blue header with a 'Menu' button on the left, the 'ARQUIVO.PT' logo in the center, and an 'Options' button on the right. Below the header is a search bar containing the text 'Ambasada RP' and a 'Search' button. A date range slider is positioned below the search bar, with '1996 1 Jan' on the left and '2021 12 Nov' on the right. Underneath the slider are three buttons: 'Pages', 'Images', and 'Narrative', along with an 'Advanced search' button on the right. The main content area shows search results. The first result is for 'Ambasada RP w Berlinie', dated '30 March 2019', with a link to 'berlin.msz.gov.pl/pl'. The second result is for 'Ambasada RP w Dublinie', dated '10 June 2017', with a link to 'dublin.msz.gov.pl/pl'. Both results include a snippet of text and a list of related terms.

Menu Options

ARQUIVO.PT

Q Ambasada RP Search

1996 1 Jan 2021 12 Nov

Pages Images Narrative Advanced search

About 23,322 results since 1996 until 2021

→ berlin.msz.gov.pl/pl

Ambasada RP w Berlinie

30 March 2019

Ambasada RP w Berlinie Masz wyłączoną obsługę języka JavaScript Polski Deutsch RSS Wersja tekstowa Być wiernym Ojczyźnie mej, Rzeczypospolitej Polskiej Aktualności **Ambasada** Współpraca dwustronna ... się zwrócić? Zasady realizacji projektów polonijnych Praktyki **Ambasada** **Ambasada** **Ambasada RP w ...**

→ dublin.msz.gov.pl/pl

Ambasada RP w Dublinie

10 June 2017

Ambasada RP w Dublinie Masz wyłączoną obsługę języka JavaScript English Polski RSS Wersja tekstowa Polsce – służyć, Europę – tworzyć, Świat – rozumieć Aktualności **Ambasada** Współpraca dwustronna ... WYDZIAŁEM KONSULARNYM **Ambasada** **Ambasada RP w Dublinie** 5 Ailesbury Road Ballsbridge D04 W221 Dublin ...

Search the **history** of a web address

Menu ARQUIVO.PT

Q bbc.co.uk Search

1996 1 Jan 2021 12 Nov

Pages Images Narrative Advanced search

Cerca de 2,538 resultados desde 1996 a 2021

Table List

2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
15 Feb	4 May	29 May	4 Jan	21 Jan	10 Jan	5 Sep	1 Jan	1 Jan	20 Jan	1 Jan	1 Jan	1 Jan	3 Jan
14 Mar	20 May	30 May	6 Jan	27 Jan	11 Jan	7 Sep	2 Jan	2 Jan	9 Feb	1 Jan	2 Jan	1 Jan	3 Jan
22 Oct	20 May	4 Aug	12 Jan	31 Jan	12 Jan	23 Sep	3 Jan	4 Jan	10 Feb	2 Jan	3 Jan	2 Jan	3 Jan
22 Oct	11 Jun		12 Jan	11 Mar	15 Jan	24 Sep	4 Jan	8 Jan	3 Apr	2 Jan	4 Jan	3 Jan	4 Jan
23 Oct	22 Jun		14 Jan	12 Mar	17 Jan	26 Sep	5 Jan	8 Jan	4 Apr	2 Jan	5 Jan	4 Jan	4 Jan
	6 Aug		21 Jan	13 Mar	18 Jan	27 Sep	6 Jan	17 Jan	5 Apr	3 Jan	6 Jan	5 Jan	4 Jan
	6 Aug		21 May	14 Mar	19 Jan	28 Sep	8 Jan	17 Jan	6 Apr	3 Jan	7 Jan	6 Jan	6 Jan
	3 Sep		1 Jul	15 Mar	22 Jan	30 Sep	10 Jan	21 Jan	7 Apr	3 Jan	8 Jan	7 Jan	6 Jan
	3 Sep		1 Dez	16 Mar	23 Jan	4 Oct	10 Jan	24 Jan	8 Apr	3 Jan	9 Jan	8 Jan	6 Jan
	4 Sep		2 Dez	17 Mar	26 Jan	5 Oct	12 Jan	28 Jan	9 Apr	3 Jan	10 Jan	9 Jan	6 Jan
	8 Sep			18 Mar	29 Jan	6 Oct	17 Jan	30 Jan	10 Apr	3 Jan	11 Jan	10 Jan	8 Jan

Browse bbc.co.uk 10 years ago

The screenshot shows the BBC website's archive interface. At the top, it says 'ARQUIVO.PT' and 'bbc.co.uk 1 July at 07:43, 2011'. The main navigation bar includes 'BBC Mobile', 'News Sport Weather Travel TV Radio More', and a search box. A left sidebar contains a 'Table' view and a date selector for 2011, with '1 July at 07:43' selected. The main content area features a 'Top News story' about Dominique Strauss-Kahn, a 'Spotlight' section on China, and a 'News' section with three articles: 'Hu warns communists on corruption', 'Chavez tells of cancerous tumour', and 'Moroccans vote on king's reforms'. A 'Most Popular in News' section is also visible at the bottom right.

Menu Options

ARQUIVO.PT


bbc.co.uk 1 July at 07:43, 2011

BBC Mobile News Sport Weather Travel TV Radio More Search the BBC

Top News story


'Doubts' over Strauss-Kahn case

The sex assault case against Dominique Strauss-Kahn appears to be in trouble amid reported concerns about the alleged victim's credibility.



Spotlight

One in a Billion





China's haves and have-nots

It is officially a Communist country - but economic data points to a stark gap between rich and poor.

Building dreams: Dai Zhikang
From hawk to billionaire: Zhou Xiaoguang
China's top art collector: Liu Yiqian

News



Moroccans vote on king's reforms

US in CIA detainee death inquiry

Poland to take over EU presidency

Crowds welcome royals in Canada

Japan manufacturers 'pessimistic'

Chavez tells of cancerous tumour

Venezuelan President Hugo Chavez says he has had a second operation to remove a cancerous tumour, in his first televised speech since flying to Cuba for treatment.

Hu warns communists on corruption

President Hu Jintao warns China's communists that corruption could cost them the support of the people, as the party marks its 90th anniversary.

Most Popular in News

Shared Read Watched/Listened

- Meet the world's most powerful computer
- Monaco Prince Albert to wed Olympic star
- China celebrates longest sea bridge

Search **images** from the past

















Menu ARQUIVO.PT Options

Q Warsaw city Search


1996 1 Jan 2021 12 Nov

Pages Images Narrative Advanced search

About 4,550,586 results since 1996 until 2021

 <p>→ dentons.com/en/fin... 5 September 2019 at 11:06</p>	 <p>→ epsmaps.com/search... 30 May 2010 at 05:50</p>	 <p>→ warsawguide.com 18 May 2016 at 05:17</p>	 <p>→ timeshighereducatio... 7 June 2019 at 02:28</p>	 <p>→ epsmaps.com/Wars... 24 December 2009 at 17:32</p>	 <p>→ timeshighereducatio... 9 December 2019 at 21:19</p>	 <p>→ iconservation.org/s... 13 May 2016 at 15:21</p>	 <p>→ realpoland.eu/packa... 14 June 2017 at 00:31</p>
 <p>→ iconservation.org/s... 13 May 2016 at 15:22</p>	 <p>→ vidiani.com/large-ro... 3 December 2019 at 11:49</p>	 <p>→ nationonline.org/o... 1 December 2019 at 00:35</p>	 <p>→ vidiani.com/detailed... 3 December 2019 at 11:44</p>	 <p>→ vidiani.com/large-de... 3 December 2019 at 11:45</p>	 <p>→ lawyerpoland.eu/ho... 31 March 2019 at 03:22</p>	 <p>→ 9flats.com/radmieci... 22 December 2019 at 11:11</p>	 <p>→ cdn.neurope.eu/nati... 28 November 2014 at 01:46</p>

ARQUIVO.PT



Visit Details

Image

Title:
Warsaw city centre

Alt text: Warsaw city centre

Caption: Poland's education system has seen some extraordinary changes in recent years. Following a huge drive in the nation's Katowice =6 1001+ Wroclaw University of Science and Technology Wroclaw Read mor...

URL:
timeshighereducation.com/sites/default/files/styles/the_breaking_news_image_style/public/warsaw_city_centre.jpg?tok=NYFWGvNj

Resolution: jpeg 620 x 413

Capture date: 7 June 2019 at 02:28

Page

Title:
Best universities in Poland | Times Higher Education (THE)

URL: timeshighereducation.com/student/best-universities/best-universities-poland

Capture date: 7 June 2019 at 02:28

Analyzing user behavior

Analyzing user behavior is an important research topic to understand users' information needs and enhance the quality of search results.



Annotation Platform

público | Público Newspaper between 1996 and 2000. Between January 1, 1996 and December 31, 2000

Very relevant

Partially relevant

Not relevant

Inaccessible content

Export

Previous

@publico.pt

http://www.publico.pt/

April 27, 1999

The screenshot shows the homepage of the Público newspaper website. At the top left, there is a navigation menu with links for 'Oracle', 'Compass', and '10 dias'. Below this, there are several sections: 'HORA TÉCNICA', 'LIVRO DE PÁTRIA', 'ENLARGES', 'JACA', 'PUBLICIDADE', 'CONSUMIDOR', 'DESTAQUE', 'SERVIÇOS', 'DOCUMENTOS', 'PUB', and 'PUBLICO'. The main content area features a blue header with 'Guia do Laser' and a date banner for 'Terça-feira, 27 de Abril de 1999'. The newspaper's logo 'Publico pt' is prominently displayed, along with a 'MAX AIR' advertisement. Below the logo, there are navigation tabs for 'NA CAPA DO JORNAL DE HOJE', 'TÍTULOS DO DIA', 'OS SUPLEMENTOS', and 'ULTIMOS DIAS'. A table lists various sections: 'ESPAÇO', 'DESPORTO', 'Sociedade', 'Educação', 'Cultura', 'Economia', 'L. Letras', 'Ultima Pátria', 'Opinião e Cartas', 'COMPTON', 'ECONOMIA', 'ARTES E CULTURA', 'SABER', 'LITERATURA', and 'PÚBLICA'. A poll asks 'Concorda com o ataque da NATO contra a Sérvia?' with 'Sim' and 'Não' options. At the bottom, there are links for 'Lazer', 'Calvin', 'Bartão', 'Vasco', 'Televisão', 'Cinecartas', and 'Fotos'.

Annotation Platform

público | Público Newspaper between 1996 and 2000. Between January 1, 1996 and December 31, 2000

Very relevant

Partially relevant

Not relevant

Inaccessible content

Export

Previous

@publico.pt <http://www.publico.pt/> April 27, 1999

Oracle
Composq
10 dias

GUIA DO LAZER

Terça-feira, 27 de Abril de 1999

Publico pt

MAX AIR

NA CAPA DO JORNAL DE HOJE

TÍTULOS DO DIA

OS SUPLEMENTOS

ULTIMOS DIAS

O Jornal de Hoje

• COMPTERQUES	Segunda-Feira
• ECONOMIA	Terça-Feira
• ARTES E CULTURA	Quarta-Feira
• SPORTS	Quinta-Feira
• OPINIÃO	Sexta-Feira
• TELEVISÃO	Sábado
• PUBLICA	Domingo

A SUA OPINIÃO

INQUÉRITO

Concorda com o ataque da NATO contra a Sérvia?

Sim Não

Lazer Calvin Bartoo Vasco Televisão Cinecartas Fotos

Rep. foto de Rep. foto de Rep. foto de

• [Ver o jornal em português](#) • [O site de serviços](#) • [Publicidade](#)

• [Links úteis](#)

Annotation Platform

público | Público Newspaper between 1996 and 2000. Between January 1, 1996 and December 31, 2000

Very relevant

Partially relevant

Not relevant

Inaccessible content

Export

Previous

@publico.pt

http://www.publico.pt/

April 27, 1999

The screenshot displays the homepage of the Público newspaper website. At the top, the date is "Terça-feira, 27 de Abril de 1999". The main logo is "Publico pt" with a globe icon. Below the logo, there are navigation tabs for "NA CAPA DO JORNAL DE HOJE", "TÍTULOS DO DIA", "OS SUPLEMENTOS", and "ULTIMOS DIAS". A table lists the sections: "COMPTÉRIOS" (Segunda-Feira), "ECONOMIA" (Terça-Feira), "ARTES E CULTURA" (Quarta-Feira), "SPORTS" (Quinta-Feira), "LITERATURA" (Sexta-Feira), and "PÚBLICA" (Sábado). A poll question asks "Concorda com o ataque da NATO contra a Sérvia?" with "Sim" and "Não" options. At the bottom, there are links for "Lazer", "Calvin", "Bartoco", "Vasco", "Televisão", "Cinecartas", and "Fotos".

Annotation Platform

público | Público Newspaper between 1996 and 2000. Between January 1, 1996 and December 31, 2000

Very relevant

Partially relevant

Not relevant

Inaccessible content

Export

Previous

@publico.pt

http://www.publico.pt/

April 27, 1999

The screenshot shows the homepage of the Público newspaper website. The left sidebar contains navigation links such as 'Oracle', 'Compos', '10 dias', and 'HORA TÉCNICA'. The main header displays the date 'Terça-feira, 27 de Abril de 1999' and the newspaper logo 'Publico pt'. Below the header, there are sections for 'NA CAPA DO JORNAL DE HOJE', 'TÍTULOS DO DIA', 'OS SUPLEMENTOS', and 'ÚLTIMOS DIAS'. A central section titled 'O Jornal de Hoje' lists various topics like 'S'África', 'Doenças Pulm', and 'Sociedade'. To the right, there is a table of contents for different sections: 'COMPTÉRIOS', 'ECONOMIA', 'ARTES E CULTURA', 'SPORTS', 'LITERATURA', and 'PÚBLICA'. Below this, there is a poll question: 'INQUÉRITO: Concorda com o ataque da NATO contra a Sérvia?' with 'Sim' and 'Não' options. At the bottom, there are links to 'Lazer', 'Calvin', 'Bartoo', 'Vasco', 'Televisão', 'Cinecartas', and 'Fotos'.

Annotation Platform

público Público Newspaper between 1996 and 2000. Between January 1, 1996 and December 31, 2000

Very relevant

Partially relevant

Not relevant

Inaccessible content

Export

Previous

@publico.pt

http://www.publico.pt/

April 27, 1999

The screenshot shows the homepage of the Público newspaper website. The left sidebar contains a navigation menu with categories like 'Oracle', 'Compass', '10 dias', and 'HORA TÉCNICA'. The main header displays the date 'Terça-feira, 27 de Abril de 1999' and the newspaper logo 'Publico pt'. Below the header, there are several sections: 'NA CAPA DO JORNAL DE HOJE', 'TÍTULOS DO DIA', 'OS SUPLEMENTOS', and 'ÚLTIMOS DIAS'. A central section titled 'O Jornal de Hoje' lists various topics. To the right, there is a table of contents for different sections like 'COMPTONERES', 'ECONOMIA', 'ARTES E SABERES', 'LITERATURAS', and 'PÚBLICA'. At the bottom, there is a poll asking 'Concorda com o ataque da NATO contra a Sérvia?' with 'Sim' and 'Não' options.

Annotation Platform

público

Público Newspaper between 1996 and 2000

Between January 1, 1996 and December 31, 2000

Very relevant

Partially relevant

Not relevant

Inaccessible content

Export

Previous

@publico.pt

http://www.publico.pt/

April 27, 1999

The screenshot shows the homepage of the Publico newspaper website. At the top, it displays the date "Terça-feira, 27 de Abril de 1999" and the newspaper's logo "Publico pt". Below the logo, there are navigation tabs for "NA CAPA DO JORNAL DE HOJE", "TÍTULOS DO DIA", "OS SUPLEMENTOS", and "ULTIMOS DIAS". The main content area includes a section titled "O Jornal de Hoje" with a list of topics such as "1ª Edição", "Debate Público", "Sociedade", "Educação", "Cultura", "Economia", "Linha", "Ultramar", and "Opinião". To the right, there is a table of contents for different sections: "COMPTON", "ECONOMIA", "ARTES E CULTURA", "SPORTS", "LITERATURA", and "PÚBLICA". Below this, there is a poll question: "Concorda com o ataque da NATO contra a Sérvia?" with "Sim" and "Não" options. At the bottom, there are several small icons for "Lazer", "Calvin", "Bartão", "Vasco", "Televisão", "Cinecartas", and "Fotos".

Annotation Platform

público | Público Newspaper between 1996 and 2000.

Between January 1, 1996 and December 31, 2000

Very relevant

Partially relevant

Not relevant

Inaccessible content

Export

Previous

@publico.pt

http://www.publico.pt/

April 27, 1999

The screenshot shows the homepage of the Público newspaper website. At the top, it displays the date 'Terça-feira, 27 de Abril de 1999' and the newspaper's logo 'Publico pt'. Below the logo, there are navigation tabs for 'NA CAPA DO JORNAL DE HOJE', 'TÍTULOS DO DIA', 'OS SUPLEMENTOS', and 'ÚLTIMOS DIÁRIOS'. The main content area includes a section for 'O Jornal de Hoje' with a list of topics such as '1ª Edição', 'Debate Público', 'Sociedade', 'Educação', 'Cultura', 'Economia', 'Linha Lenta', and 'Última Página'. To the right, there is a table of contents for the day's sections: 'COMPTONERES', 'ECONOMIA', 'ARTES E CULTURA', 'SPORTS', 'LITERATURA', and 'PÚBLICA'. Below this, there is a poll titled 'INQUÉRITO' asking 'Concorda com o ataque da NATO contra a Sérvia?' with 'Sim' and 'Não' options. At the bottom, there are links to 'Lazer', 'Calvin', 'Bartão', 'Vasco', 'Televisão', 'Cinecartas', and 'Fotos'. The left sidebar contains various links and information, including 'Oracle', 'Compass', '10 dias', and 'PÚBLICO é um site associado'.

Annotation Platform

público | Público Newspaper between 1996 and 2000. Between January 1, 1996 and December 31, 2000

Very relevant

Partially relevant

Not relevant

Inaccessible content

Export

Previous

@publico.pt

http://www.publico.pt/

April 27, 1999

The screenshot shows the homepage of the Público newspaper website. The left sidebar contains a navigation menu with categories like 'Oracle', 'Compass', '10 dias', and 'HORA TÉCNICA'. The main header displays the date 'Terça-feira, 27 de Abril de 1999' and the newspaper logo 'Publico pt'. Below the header, there are navigation tabs for 'NA CAPA DO JORNAL DE HOJE', 'TÍTULOS DO DIA', 'OS SUPLEMENTOS', and 'ÚLTIMOS DIAS'. The central content area includes a section for 'O Jornal de Hoje' with a list of topics such as 'ISRAEL', 'Doutor Pólvora', and 'Sociedade'. To the right, there is a table of contents for different sections like 'COMPTONERES', 'ECONOMIA', 'ARTES E SABERES', 'SAÚDE', 'LITERATURA', and 'PÚBLICA'. At the bottom, there is a poll titled 'INQUÉRITO' asking for opinion on NATO's attack on Serbia, with 'Sim' and 'Não' options.

Annotation Platform

público | Público Newspaper between 1996 and 2000. Between January 1, 1996 and December 31, 2000

Very relevant

Partially relevant

Not relevant

Inaccessible content

Export

Previous

@publico.pt

http://www.publico.pt/

April 27, 1999

The screenshot shows the homepage of the Público newspaper website. At the top, it displays the date 'Terça-feira, 27 de Abril de 1999' and the newspaper's logo 'Publico pt'. Below the logo, there are navigation tabs for 'NA CAPA DO JORNAL DE HOJE', 'TÍTULOS DO DIA', 'OS SUPLEMENTOS', and 'ÚLTIMOS DIAS'. The main content area is divided into several sections, including 'O Jornal de Hoje', 'A SUA OPINIÃO', and a poll asking 'Concorda com o ataque da NATO contra a Sérvia?'. The poll has two options: 'Sim' and 'Não', with 'Ver' buttons below each. At the bottom, there are links to various sections like 'Lazer', 'Calvin', 'Bartoo', 'Vasco', 'Televisão', 'Cinecartas', and 'Fotos'.

Annotation Workflow

público | Público Newspaper between 1996 and 2000. Between January 1, 1996 and December 31, 2000

Very relevant Partially relevant Not relevant Inaccessible content Export Previous

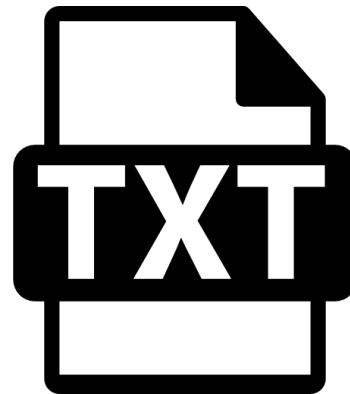
@publico.pt http://www.publico.pt/ April 27, 1999



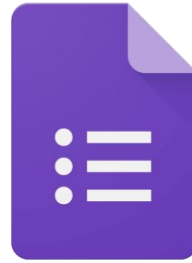
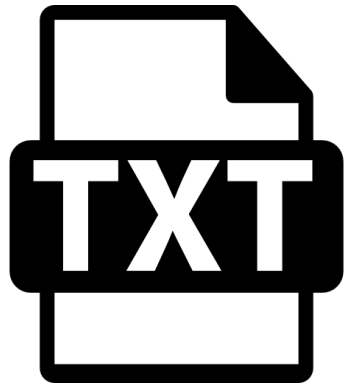
público | Público Newspaper between 1996 and 2000. Between January 1, 1996 and December 31, 2000

Very relevant Partially relevant Not relevant Inaccessible content Export Previous

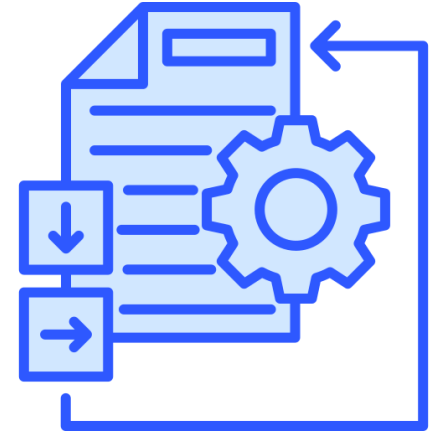
@publico.pt http://www.publico.pt/ April 27, 1999



Annotation Workflow



Google Forms



Annotation Results and Analysis

Metric	Search Engine Production	New Search Engine (SOLR)
Mean Average Precision	0.3582	0.3147
Precision@1	0.78	0.48
Precision@5	0.62	0.54
Precision@10	0.56	0.51
Precision@100	0.10	0.09
nDCG@10	0.55	0.46

Future Work

- Use this data to further tune Solr field weights and explore alternative scoring models (e.g., BM25 parameter tuning, learning-to-rank approaches).



Thank You

contacto@arquivo.pt

