

Enhancing access to research the Geocities historical collection.

Pedro.Gomes@fccn.pt

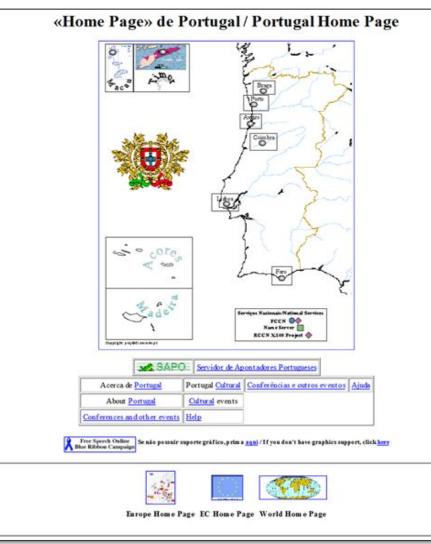
Daniel.Gomes@fccn.pt





CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

ARQUIVO.PT Present Arquivo.pt



CERN European Laboratory for Particle Physics

Lab - News - Activities - Physics - Other Subjects - Search - Shrink - Expand

Welcome to the European Laboratory for Particle Physics, located near Geneva in Switzerland and France.

The <u>WWW support team</u> provides a set of <u>Services</u> to the physics experiments and the lab.

About the Laboratory

General:

<u>Help</u> and <u>General information</u>, <u>Divisions</u>, <u>groups and activities</u> (structure), <u>Scientific Committees</u>, <u>Vi</u> experiments, <u>Education Services</u>, <u>Recruitment and Vacancies</u>.

Databases:

Directories (phone & email, services & people), <u>Buildings</u> - <u>Scientific Information Service</u> (library, a **Agenda and News**:

Press Releases, News from the Users' Office, current Seminars, CERN Schools & Conferences, intern news.

Hot News

<u>1996 European School of High Energy Physics</u>, Carry-le-Rouet (near Marseille), France <u>1996 CERN School of Computing</u>, Egmond aan Zee, the Netherlands, 8-21 September, 1996. <u>Anti-Hydrogen</u>, First Atoms of Antimatter produced at CERN <u>Linac96</u>, XVIII International Linac Conference, CERN, Geneva <u>Older News</u>, Hot News (older items)

http://s700.uminho.pt/homepage-pt.html - 1996

http://www.cern.ch - 1996



Arquivo.pt preserves national and international historical web content



CERN European Laboratory for Particle Physics

Lab - News - Activities - Physics - Other Subjects - Search - Shrink - Expand

Welcome to the European Laboratory for Particle Physics, located near Geneva in Switzerland and France.

The <u>WWW support team</u> provides a set of <u>Services</u> to the physics experiments and the lab.

About the Laboratory

General:

<u>Help</u> and <u>General information</u>, <u>Divisions</u>, <u>groups and activities</u> (structure), <u>Scientific Committees</u>, <u>Vi</u> experiments, <u>Education Services</u>, <u>Recruitment and Vacancies</u>.

Databases:

<u>Directories</u> (phone & email, services & people), <u>Buildings</u> - <u>Scientific Information Service</u> (library, a **Agenda and News**:

Press Releases, News from the Users' Office, current Seminars, CERN Schools & Conferences, intern news.

Hot News

<u>1996 European School of High Energy Physics</u>, Carry-le-Rouet (near Marseille), France
 <u>1996 CERN School of Computing</u>, Egmond aan Zee, the Netherlands, 8-21 September, 1996.
 <u>Anti-Hydrogen</u>, First Atoms of Antimatter produced at CERN
 <u>Linac96</u>, XVIII International Linac Conference, CERN, Geneva
 <u>Older News</u>, Hot News (older items)



ARQUIVOPT Keep improving Full-text search engine & User interfaces

Menu	ARQUIVO.PT	Options
	Q Coronavirus Search	
	1996 1 Jan 24May	
	Pages Images Advanced search	
	About 293,653 results since 1996 until 2021	
	Coronavirus	
	8 December 2017 Coronavirus Deutsch English Русский Über uns Presse Impressum Sitemap Kontakt Mega-Menu ist? HOME » CENTER » Erregersuche A- Z » Coronavirus Erregersuche A-Z E-Learning-Tool Anrufer aus Österreich: Tel.: 02236 646 3070 Coronavirus (inkl. SARS- und MERS-CoV) (behülltes Virus	
	→ bode-science-center.pt/center/erregersuche-a-z/coronavirus.html Coronavirus	
	17 July 2018 Coronavirus Deutsch English Русский Über uns Presse Impressum Sitemap Kontakt Mega-Menu ist? HOME » CENTER » Erregersuche A-Z » Coronavirus Erregersuche A-Z E-Learning-Tool Anrufer aus Österreich: Tel.: 02236 646 3070 Coronavirus (inkl. SARS- und MERS-CoV) (behülltes Virus	
	View more results of bode-science-center.pt	
	→ bode-science-center.com/center/relevant-pathogens-from-a-z/coronavirus.html	
	Coronavirus	
	11 April 2019	
	Coronavirus Deutsch English Русский About us Press Corporate Information Sitemap hand hygiene HOME » CENTER » Relevant pathogens from A-Z » Coronavirus Relevant pathogens from A questions please send an e-mail to BODE SCIENCE CENTER. Coronavirus (incl. SARS- and MERS-CoV	

To launch new services

ARQUIVOPT Like the Images Searching & High Quality Archiving Services

Menu			ARQU	IVO.PT			Options
		Q Coronavirus			Search 2021 24May Advanced search		
	encia.com/saude/casos 4 February 2020 at 18:40	- thenakedscientists.com/po 6 March 2020 at 01:30	→ unicamp.br/unicamp/notici 12 March 2020 at 18:23	→ portaldaenfermagem.com 11 February 2020 at 20:47	• medlineplus.gov/coronavir 13 July 2018 at 03:10	→ rtve.es/noticias 12 March 2020 at 16:04	→ phmk.es/politica-sanitaria 3 March 2020 at 12:20
→ farodevigo.es:80/espana/2 7 March 2020 at 02:13 → WW	ONAVIRUS STRUCTURE AND GENE EXPRESSION	NOVO I NEW 新型 短状病毒 CORONAVÍRUS 2019 .c.v ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓	→ portaldaenfermagem.com 11 February 2020 at 20:41	→ elnacional.cat/es/salud/si 11 February 2020 at 23:17	Coronavirus → ec.europa.eu/portugal 16 March 2020 at 19:41	→ liberoquotidiano.it/news/li 8 February 2020 at 15:37	→ elnacional.cat/es/salud.ht 24 February 2020 at 12:54

And at the same time ... ARQUIVO.PT

Increase our collection and users community





- Geocities.com was founded in 1994, and in 1999, Geocities was acquired by Yahoo! and it became the third most popular platform;
- In April 2009, Yahoo!, announced that it would shut down Geocities, deleting more than 38 million pages created by users all around the world;
- The Archive team announced a project to archive Geocities pages, resulting in a torrent file;



ARQUIVO.PT Initiatives

http://www.oocities.org/

A living memorial on early Web culture and an effort to maintain an amazingly extensive co

Welcome to OoCities.org

In October 2009 we archived our olden cities of the web: The unique pages on "GeoCities.com" just before they were taken down. Geocities started in 1994 and was the first giant "social network" and one of the most important websites in the world until it was closed on 27.10.2009. Our aim is to save those pages which are worthy and unique scientific sources or are of great public interest as well as those, which are historically interesting or just representing the 90's website culture and style. If you are interested in these old pages feel free to browse around. Please note: it will be a while until all pages are sorted, complete and free of any spam contents! In case you are searching for your own personal site, there is a small chance that we have it but have not put it online yet. In case you want your own page, email or picture to be removed, contact us as well. For any questions, read our FAQ or contact us at *oocities {AT} gmail {DOT} com*

GeoCities Neighborhoods

Until 1999, GeoCities pages were sorted into categories using the following scheme: GeoCities.com/neighborhood/number starting with number 1000 and counting up to 9999. Thus each neighborhood (=category) or subneighborhood contains up to 9000 pages. (Full list of GeoCities Subneighborhoods)

Area51 ::: Science ficition & fantasy Athens ... Teaching, education, philosophy, reading, writing Augusta ::: Golf Baja ::: Off-road & adventure travel BourbonStreet ...: Jazz, New Orleans, Cajun food, Southern United States topics CapeCanaveral ::: Science, mathematics, aviation, technology, engineering, CapitolHill ...: Government and Politics CollegePark ::: University life Colosseum ::: Athletics & sports EnchantedForest ::: Topics of interest to children Hollywood ::: Films and actors MotorCity ::: Automobiles & racing NapaValley ::: Wine Nashville ::: Country music Petsburgh ::: Pets Pipeline ::: Extreme sports RainForest ::: Conservation

Oocities Like Page Docities hå cerca de um ano https://medium.com/404-

Special Thanks goes to... Special Thanks goes to all the editors and webmasters who made this project alive by updating their old links every day from * geocities.com/*

to *.00cities.com/* : for instance those at



Team members wanted There's still just a small group of people working on OoCities, but there is a lot of work to get done. Please contact us if won're interested

ENHANCED BY Google

If you remember an old

geocities username exactly,

then you can just try to go to

"** oocities.com/**username"

without searching or looking

(The google search above

geocities.ws, reocities.com,

deletedcity.net. Google is not

aware of everything avaiable

web.archive.org which is not

but major parts except for

internetarchaeology.org,

includes results from

oocities.com/org,

webcitation.org,

in google.)

Search

the website

through the lists.

Search

Custom Search from Google

A living memorial on early Web culture and an effort to maintain an a

Todos os resultados oocities

Aproximadamente 366 resultados (0.21 segundos)

Exibindo resultados para *Cape Canaveral* Em vez disso, pesquisar por *CapeCanaveral*

Cape Canaveral-Cocoa Beach Historical Trail

www.geocities.ws > krdvry > hikeplans > cp_canaveral-cbch > plancapecana Cape Canaveral was renamed as Cape Kennedy because of that president's intere his ...

Stipe's company to film Coupland's new book: Rock star sitcom for .. www.oocities.org > soho > gallery > npost3

To be published in September by Bloomsbury Press, the **Cape Canaveral**-set All Fai before the ...

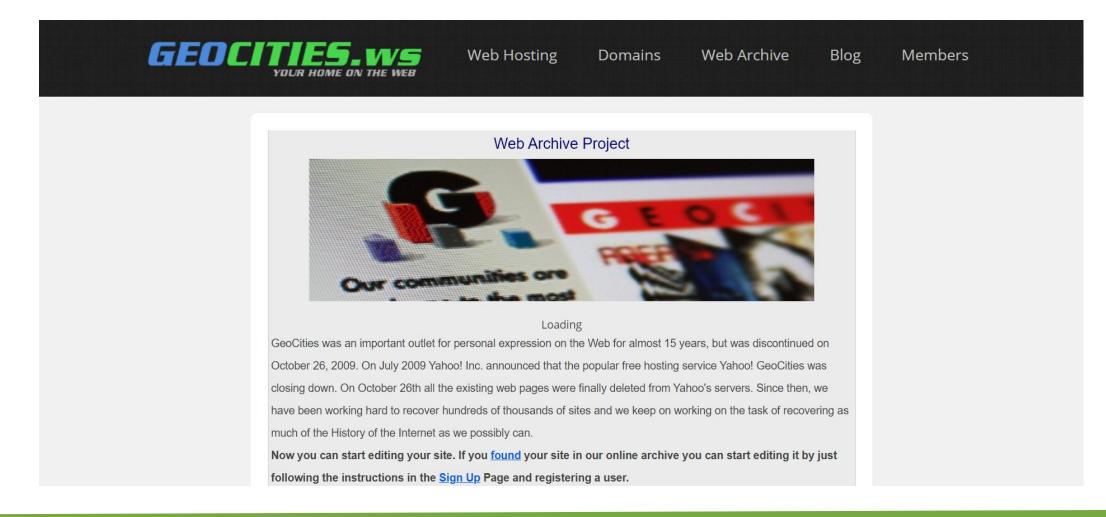
Marcado oocities







http://www.geocities.ws/archive/



ARQUIVOPT Geocities Arquivo.pt

Menu	ARQUIVO.PT	Options
	Q virus collection:Geocities Search	
	1996 1 Jan	
	1 Jan 28 May Pages Images Advanced search	
	About 45,908 results since 1996 until 2021	
	\rightarrow de.geocities.com/franky_hbs/virus.htm	
	virus	
	29 October 2009 virus VIRUSWARNUNG Codename des Virus ist "ARBEIT". Wenn du "Arbeit" von irgendwo bekommst, ob von deinen Kollegen oder deinem Chef, per E-Mail oder via Internet, öffne es nicht, schaue es nicht an Bekannten hast, bedeutet dies, dass Du infiziert bist und der Virus "Arbeit" dein Leben schon	
	→ geocities.com/Pipeline/Halfpipe/6151/Hacker/virus.htm	
	Virus	
	18 October 2009	

ARQUIVO.PT Geocities Arquivo.pt

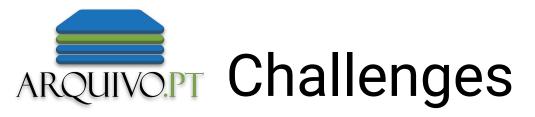
Menu	ARQUIVO.PT				Options
Q virt	us collection:Geocities			Sea	rch
1996 1 Jan	•	•	2021 28May		
Pages	Images			Advance	d search
About 1,6	688 results since 1996 until 2021				
→ geocities.com/jaapsch/p 27 April 2009 at 17:18	→ geocities.com/jaapsch/p 27 April 2009 at 17:18	→ nlm.nih.gov/medlineplus 28 June 2009 at 01:51	Hemorrhage tab appears by by by by by by by by by by	→ geocities.com/medipedi 26 October 2009 at 21:53	→ geocities.com/medipedi 26 October 2009 at 21:53

ARQUIVO.PT Geocities Arquivo.pt

Menu		ARQUIVO.PT	Options
		geocities.com/1000agres 26 February at 01:50, 2008	
III Table			
2008	۲		
February	۲	Bem vindo à Página Oficial dos 1000Agres!	
26 February at	01:50	1000Agres é uma banda formada há cerca de 3	
2009	• 🕞	anos que tem como objectivo o estudo da relação da música portuguesa com a música africana.	
2014	©	1000Agres são quatro músicos que vos propõem um espectáculo de cerca de duas horas que intitularam *1000Agres do Mar ^a e que consiste numa original	
2015		abordagem aos diversos temas do património musical destas duas culturas.	
2016		Esse mar que nos separa mas que ao mesmo tempo foi a estrada para o encontro.	
2017	O	África, que nos influenciou e que foi influenciada, conta-nos tantas histórias de rara beleza que nos fascina interiormente e Portugal que ao dar novos	
2018		mundos ao mundo, foi construindo o seu património musical tão imensamente variado num país geograficamente tão pequeno.	
		Visitante N.* Visitante N.* Visitante N.* Visitante N.* FastCounter by LinkExchange Do Linal 1000agres@netvisao.pt	

ARQUIVO.PT Motivation

Menu		ARQUIVO.PT Options
		g <u>eocities.com:80</u> 13 October at 21:37, 1996
I Table		The ? is the set
1996	۲	Welcome to <u>GeoCities</u> Your Home On The Web!
October	۲	What does this symbol mean?
13 October at 21	:37	Today's <u>Cool</u> Home Page:
2008	۲	SoHo 1085 - Mike's got a Collection of Animated GIFs that must be experienced!
2009	•	Enter Geo Cittes The Here!
2010	۲	Tell us what you think!
2013	۲	Take the first <u>GeoCities User Survey</u> !
2014		TAKE
2015	۲	STOCK!
2016	۲	Enter for a chance to win \$2500 worth of stock in your favorite Internet company.
2017	۲	[New Stuff] [Info] [Free Homepages] [FAQ] [Announcements] [Neighborhoods] [Cool] [Mail] [Business_Partners] [Cameras] [Search] [Utilities] [Bulletin_Board] [Neighborhood_Watch] [Help] [Map]
2018		© 1995,1996 GeoCities. All rights reserved

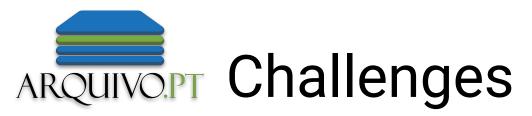




webrecorder/warcit

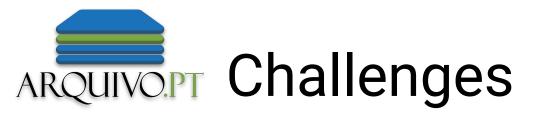
Convert Directories, Files and ZIP Files to Web Archives (WARC)



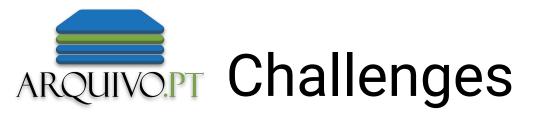


Choose the URL default for each folder:

- http://www.geocities.com vs http://geocities.com
- http://uk.geocities.com vs http://www.geocities.com



$\overleftarrow{\bullet}$ > C $\widehat{\bullet}$	💿 🔒 https://arquivo.pt/wayback/20091026192022/http://geocitie	es.com/a1kc_radio/
Menu	ARQUIVO.P	Т
III Table	AR	QUIVO.PT
	Sear	rch in other archives
	\leftarrow \rightarrow C \textcircled{a}	💿 🖴 https://arquivo.pt/wayback/20091026192022/http://uk.geocities.com/a1kc_radio/ 🚥 😒 🏠
	Menu	ARQUIVO.PT
		uk.geocities.com/a1kc_radio 26 October at 19:20, 2009
	III Table	
	2009 💽	A1kC Productions
	October 🔺	
	26 October at 19:20	Radio Jackie North
	26 October at 19:20	The Un-Official Radio Jackie North Web Site



. Identify the correct timestamp for each webpage





- 1. Unzip all files. Examples:
 - a. Lowercase \rightarrow YAHOOIDS \rightarrow a \rightarrow a \rightarrow aa_chan01
 - b. Subsites \rightarrow YAHOOIDS \rightarrow geocities \rightarrow uk.geocities.com \rightarrow a_arif_ali
- 2. For each subfile (e.g., aa_chan01):
 - a. Get the new timestamp from the last update of the files inside in each folder (e.g., 2009 11 05).
 - b. warcit --fixed-dt timestamp http://www.geocities.com/ path_folder.
 - If there is a subsite, "http://www.geocities.com/" will be replace by "http://uk.geocities.com/".
- 3. Merge each individual WARC file and removing the "revisit" records since will have the current timestamp (i.e., 2021).



arquivo.pt/searchGeocities

- . full text page search
- image search
- page replay



Insert keyword or URL

Search pages from the past

About this project

Português



- From: 2009-04-24
- Until: 2010-10-13
- Disk space: 280 GB (war
- 35 millions collected file

rc.gz)			<page-header></page-header>	
es	Geocities homepage in 1996	Geocities homepage in 2008	Geocities homepage in 2009	Country & WesternBeauties
			Exercise Constraints of the second seco	Officience (Floregour Chargebour
	Queen – Manu's Musica	Artbook Online Gallery – Collage + digital art	Computers	Ilustrious Portugues Navigators
	Freedown of the free of the fr		Cour Hongs	
	Unofficial F3000	Midis Internacionais (music files in midi)	Our Hommer. All about dogs	My personal diary

Portuguese

Examples of Geocities preserved websites



Menu	ARQUIVO.PT	Options
	ar.geocities.com/manusmusic/queen.htm 10 November at 01:58, 2009	
I Table		
2009		
November	No se por qué quieren suponer que	
10 November at 01:58	101:58 nos inspiramos en la música de otras bandas. Queen es demasiado grande para eso" Fredie Mercury 1986.	
	 Obiscografia Obiscografia Obiscografia Obiscografia Oconología Obiscografia Obiscografia<th></th>	



Menu	ARQUIVO.PT	Options			
	de.geocities.com/artboook2005/collage.htm 26 October at 21:41, 2009				
I Table	. ARTBOOOK ONLINE GALLERY .				
	STAR SHOP AMERICA offers USA motifs on T-Shirts, Sweat-Shirts, Coffee Cups, Mugs, Posters, Postcards, Greeting Cards, Stamps, Buttons, Stickers, Journals, Mousepads, Caps, Infant Wear, Tile Boxes & much more - We have all kinds of Art related products in store ! - Visit our VISUAL ARTS SECTION with Paintings, Collages, Pho	<u>tos, Cartoons, Posters & more !</u>			
2009					
October 🍝	<u>COLLAGE + Digital Art</u>				
26 October at 21:41	Surrealistic collages - Digitally produced artworks				
	collage, collage > digital > collage exhibits + digital art galleries > Digital Art > Surrealism > Fantasy Art				
	collage soligitat > collage exhibits + digital att galleries > Digital Art > Surealism > Expension + Digital Art > Surealism > Digital Art > Digital Art > Surealism > Digital Art > Digit				
	COLLACE				
	<u>COLLAGE</u>				



Menu		ARQUIVO.PT	Options
		<u>c.geocities.com/peterochocki/computers/1980comp/appmac.htm</u> 20 October at 10:35, 2009	
I Table	Back to Home Page		Help Notes
2009			
October 💽		The Apple Macintosh	
20 October at 10:35			
	Manufacturer	Apple Computers (US)	
	Model	Macintosh (From a variety of apple.)	
	Date Launched	January 1984	
	Price	£1600	
	Microprocessor type	Motorola 68000 @ 8 MHz	
	ROM size	64 kilobytes	
	Standard RAM	128 kilobytes	
	Maximum RAM	128 kilobytes	
	Keyboard type	Typewriter style	







arquivo.pt/api

 Available for automatic processing

APIs

- . CDX-server API
- . Memento API
- Page Search API
- Image Search API

```
G
                ി
                                         https://arquivo.pt/imagesearch?g=gueen&more=safe.imgThumbnailBase64&prettyPrint=true&collection
        Raw Data
                    Headers
JSON
Save Copy Pretty Print
  "serviceName": "Arquivo.pt - image search service.",
 "linkToService": "https://arquivo.pt/images.jsp",
 "linkToDocumentation": "https://github.com/arquivo/pwa-technologies/wiki/ImageSearch-API-v1.1-(beta)",
 "linkToMoreFields": "https://arquivo.pt/imagesearch?q=queen&prettyPrint=true&collection=Geocities&more=pageHost,matchingImages,safe",
 "nextPage": "https://arquivo.pt/imagesearch?q=queen&more=safe,imgThumbnailBase64&prettyPrint=true&collection=Geocities&offset=50",
 "previousPage": "https://arquivo.pt/imagesearch?q=queen&more=safe,imgThumbnailBase64&prettyPrint=true&collection=Geocities&offset=0",
 "totalItems": 17032,
 "numberOfResponseItems": 50,
 "offset": 0,
  "responseItems": [
      "pageTstamp": "20101020064855",
     "safe": 0.9681148,
     "imgWidth": 144.
      "imgMimeType": "image/jpeg",
      "imgDigest": "e25424bf0ec9434111ea0ff339ef56bb23d812a3b51855a5a38408c82f2a31b2",
      "imgTstamp": "20101020064855",
     "imgSrc": "http://www.geocities.com/detroitrockandroll/photos1/queenbee.jpg",
      "pageURL": "http://www.geocities.com/detroitrockandroll/bands5.html",
      "pageTitle": "Detroit Rock and Roll - P-R",
      "imgHeight": 144,
      "imgCaption": [
        "OUEEN BEE"
      "imgAlt": [
        "Oueen Bee"
      "imgTitle":
        "Queen Bee"
      "collection": [
        "Geocities"
      "imgLinkToArchive": "https://arquivo.pt/wayback/20101020064855im_/http://www.geocities.com/detroitrockandroll/photos1/queenbee.jpg",
      "pageLinkToArchive": "https://arquivo.pt/wayback/20101020064855/http://www.geocities.com/detroitrockandroll/bands5.html"
      "pageTstamp": "20091028232513",
     "safe": 0.99945766,
      "imgWidth": 520,
      "imgMimeType": "image/gif",
      "imgDigest": "6f5fb648e4b44ad681dbc411012985e453488c82948051373b3b5d76c1ecbbc0",
      "imgTstamp": "20091028232513",
     "imgSrc": "http://de.geocities.com/elite abi2004/assets/images/autogen/The Queen NBanner.gif",
      "pageURL": "http://de.geocities.com/elite_abi2004/html/thequeen.html",
      "pageTitle": "TheQueen",
      "imgHeight": 75,
      "imgCaption": [
       "Frau Alfes - The Queen Dagmar Alfes, von manchen heimlich "Daggi" genannt, kann man einfach nicht verfehlen. Sie ist die\nTHE QUEEN... von den PAP
      "imgAlt": [
        "The Oueen"
      "imgTitle": [
       "The Queen"
      "collection": [
        "Geocities"
     "imgLinkToArchive": "https://arquivo.pt/wayback/20091028232513im_/http://de.geocities.com/elite_abi2004/assets/images/autogen/The_Queen_NBanner.gif",
      "pageLinkToArchive": "https://arquivo.pt/wayback/20091028232513/http://de.geocities.com/elite_abi2004/html/thequeen.html"
```



- Originate innovative scientific contributions.
- National and international researchers could benefit from the access tools of Arquivo.pt.
- This process can be applied easily to any kind of collection.



- **Special Collection** \rightarrow Coronavirus COVID-19 outbreak
 - IIPC Collaborative Collection
 - Automatically queried a web search engine
 - e.g., "coronavirus site:news-site.pt"
 - Human review
- New Service Save Page Now
- New Version of Image Search
 - Search among 1.8 billion images archived since 1994
 - Image captions extracted from text close to images in HTML





• 3rd place: 2 000 €



More information

arquivo.pt/searchGeocities > About this project

Português



Insert keyword or URL

Q

Search pages from the past

About this project