

Págico: Evaluating Wikipedia-based information retrieval in Portuguese

Cristina Motta (Linguateca/FCCN)

Alberto Simões (CEHUM/UM)

Cláudia Freitas (PUC-Rio / Linguateca)

Luis Costa (Linguateca/FCCN)

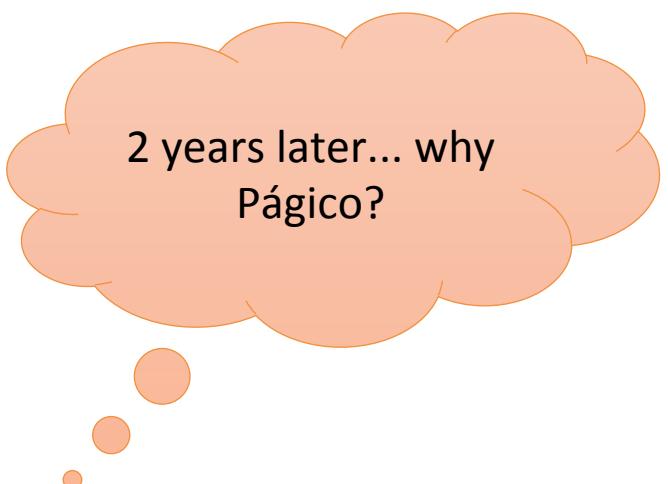
Diana Santos (Univ. of Oslo/Linguateca)



Págico – Português Mágico



- Evaluation contest
- Organized by Linguateca, a distributed language resource center for Portuguese
 - <http://www.linguateca.pt>
- Announced in June 2011
- Results delivered in the beginning of 2012





Págico – Português Mágico

- Evaluation contest
 - Systems (and humans) must find non-trivial answers to complex information needs in Portuguese Wikipedia (150 topics were created):
 - Which Portuguese-speaking football players played professionally in more than three different countries?
 - Jornais que circularam no Rio de Janeiro entre 1910 e 1960
 - Produtos agrícolas com os quais se pode produzir combustível em escala comercial
 - Nomes ligados à luta contra o racismo no sec XX no Brasil
 - Filmes brasileiros premiados na categoria Montagem
 - Pratos brasileiros de origem ou influência indígena

answer to information needs that would require browsing hundreds of pages

“we are **not proposing** evaluation of toy systems just to check if **computers can be as good as humans**. Rather, we are **interested in systems that really help humans** in non trivial information³ finding



Págico – Português Mágico

- The task:
 - Find (PT)Wikipedia pages that answer the topics. The answer must be **the name of the page**.
 - Additional pages must be provided when the answer page is incomplete
 - Pages (answers and justifications) must be selected from a static version of the portuguese Wikipedia.
 - **Pagico Collection:** a Wikipedia snapshot of April 25th 2011
 - Documents converted in XHTML files
 - Without *Infoboxes*



The Collection: Wikipedia folksonomy

- Distribution of categories across PT Wikipedia pages
 - 99.446 categories classify 681.058 documents

# Documents	# Categories	Percent
]0, 1]	32 652	34.21%
]1, 66]	59 775	62.63%
]66, 130]	1 789	1.87%
]130, 194]	507	0.53%
]194, 260]	231	0.24%
]260, 345]	166	0.17%
]345, 442]	108	0.11%
]442, 592]	84	0.09%
]592, 862]	68	0.07%
]862, ∞[65	0.07%

# Categories	# Documents	Percent
0	8 771	1.271%
]0, 8]	676 705	98.097%
]8, 15]	4 008	0.581%
]15, 23]	314	0.046%
]23, 33]	25	0.004%
]33, ∞[6	0.001%

Table 8: Number of categories per document.

Table 7: Number of documents per category number.



Topic creation

- Topics should be interesting to the Portuguese-speaking community
- The answers
 - should not be obvious
 - ...and not obvious to Google(!), in order to justify a wikipedia search
 - should be spread among different pages in PT Wikipedia
- Topic: Músicos do início da Bossa Nova / *Musicians associated with the development of Bossa Nova*
 - Browse Bossa_Nova page..

It doesn't involve complex search...
...from the standpoint of humans.

Topic creation

- PT Wikipedia in 150 topics
 - 8 Super-themes
 - 25 subthemes
 - One or more areas/subareas were assigned to each topic:
 - “*Documentary films about brazilian politicians*”
 - Theme: Artes; Letras; Política
 - Subtheme: cinema; política; história
 - “*Comidas de santo (comidas rituais do Candomblé ou Umbanda) que também fazem parte da culinária brasileira*”// *religious food that are also brazilian food*
 - Theme: Cultura
 - Subtheme: culinária; religião; antropologia; folclore

Tema -subtema	#	Tópicos %
Letras	69	46,00
- história	50	33,33
- literatura	15	10,00
- linguística	6	4,00
- jornalismo	3	2,00
- filosofia	2	1,33
Artes	36	24,00
- música	19	12,67
- cinema	10	6,67
- televisão	4	2,67
- artes plásticas	2	1,33
- artes	2	1,33
Geografia	34	22,67
- geografia	26	17,33
- arquitetura/urbanismo	7	4,67
- demografia	4	2,67
- geologia	2	1,33
Cultura	27	18,00
- antropologia/folclore	12	8,00
- religião	7	4,67
- culinária	5	3,33
- cultura	3	2,00
- ensino	2	1,33
Política	19	12,67
Desporto/Espor tes	18	12,00
Ciência	14	9,33
- saúde	4	2,67
- zoologia	3	2,00
- ciência	2	1,33
- botânica	2	1,33
- geologia	2	1,33
- matemática	1	0,67
Economia	6	4,00





Answer documents

- A correct answer , according to Págico:
 - Wiki-PT page title → expected semantic class

“Gêneros musicais que misturam samba
e gêneros norte americanos” → “Turma_da_Pilantragem” X

“Que países têm amarelo na bandeira?” → “Bandeira_do_Brasil” X

Good answer from a practical perspective...

Bad answer from a logic perspective

Correct answer

X

Useful answer

.. how to decide/define what should be considered a “good answer” ?



Answer documents – justification pages

- Additionally, participants need to provide the wikipedia pages that support that the chosen answer is, indeed, the correct one

Pagico_103
Tema(s): Cultura, Geografia
Dados geográficos: Brasil

[Página principal](#) | [Sair](#) [observador_hum]



Movimentos culturais surgidos no **nordeste** do Brasil.

Do we need justification here?

Tema: Qualquer tema

Ir para página visualizada anteriormente

Voltar às respostas a este tópico

Tópico Anterior Próximo Tópico

Lista de Tópicos

Manguebeat

Manguebeat (também grafado como manguebit ou mangue beat) é um movimento musical que surgiu no [Brasil](#) na [década de 90](#) em [Recife](#) que mistura ritmos regionais, como o [maracatu](#), com [rock](#), [hip hop](#), [samba](#), [reggae](#) e [música eletrônica](#). Esse estilo tem como ícone o músico [Chico Science](#), que também era ambientalista e divulgador das idéias, ritmos e contestações da ecologia. Chico Science é considerado o fundador do [Movimento Manguebeat](#) e autor do primeiro manifesto do Movimento, intitulado "Manguebeat". O objetivo do movimento surgiu de uma metafísica de resistência ao capitalismo globalizado. O Manguebeat precisava ser uma forma de expressão artística que pudesse combinar a diversidade, a agitação na música contaminada por ritmos regionais e a busca por novas formas de vida. Com o surgimento de várias bandas no cenário musical brasileiro, o Manguebeat se tornou uma referência importante para o rock brasileiro. As bandas mais conhecidas do movimento são [Pernambuco](#) e [Brasil](#), sendo o principal representante do movimento. Com o surgimento de várias bandas no cenário musical brasileiro, o Manguebeat se tornou uma referência importante para o rock brasileiro.

Págico: only systems must provide justification

LogOnto 2014 - FGV - Brasil



Answer documents – justification pages

Cantores **vaiados** nos festivais de música brasileira na década de 60.

Tema: Qualquer tema

Tópico Anterior Próximo Tópico

[Ir para página visualizada anteriormente](#)

[Voltar às respostas a este tópico](#)

Festivais de MPB nos anos de 1960

No festival de [1967](#) faria sucesso também com [Roda Viva](#), interpretada por ele e pelo grupo [MPB-4](#) — amigos e intérpretes de muitas de suas canções. Em [1968](#) voltou a vencer outro *Festival*, o III [Festival Internacional da Canção da TV Globo](#). Como compositor, em parceira com [Tom Jobim](#), com a canção *Sabiá*. **Mas desta vez a vitória foi contestada pelo público**, que preferiu a canção que ficou em segundo lugar: [Pra não dizer que não falei de flores](#), de [Geraldo Vandré](#).

A participação no *Festival*, com *A Banda*, marcou a primeira aparição pública de grande repercussão apresentando um estilo amparado no movimento musical urbano carioca da [Bossa nova](#), surgido em [1957](#). Ao longo da carreira, o samba e a MPB também seriam estilos amplamente explorados.

Trilha-sonora e adaptações de livros

[thumbnail|200px|esquerda|Chico](#) participou como autor e compôs várias canções de sucesso para o filme *Quando o*

To be contested by the audience = to be booed?

Answers in Págico

Topics with most and least correct answer documents

Topic	#
Indigenous tribes living in the Amazon Rainforest	95
Museums in capitals of Lusophone countries	62
Locations mentioned in "Os Lusíadas"	51
Indigenous Brazilian peoples considered extinct.	50
Viceroyes of the Portuguese India	48
...	
Politicians from Portuguese speaking African countries who studied in the Soviet Union	2
Churches in Rio de Janeiro constructed by Afro-Brazilian confraternities.	1
Members of Parliament from FRELIMO	1
Mozambican writers who received Prémio Camões	1
Foreign writers who visited Portugal in the 19th Century and published descriptions of their travels	1

The most correct topics

	ID	Topic	Total	Hum	Sys	H & S
H	135	Aves de Angola	54	10	44	0
	19	Tribos indígenas que vivem na Amazônia.	115	56	35	24
	90	Filmes brasileiros premiados na categoria Montagem.	34	8	19	7
	13	Dinossauros carnívoros que habitaram o Brasil.	23	6	12	5
S	19	Tribos indígenas que vivem na Amazônia.	115	56	35	24
	62	Praias de Portugal boas para a prática de surf	30	5	6	19
	7	Guitarristas portugueses que também foram compositores.	34	17	0	17
	11	Filmes sobre o cangaço.	41	20	4	11



Págico – by-products

- CARTOLA – <http://www.linguateca.pt/Cartola>
 - Freely available package
 - Págico Collection: 681.058 documents from wikipedia portuguesa de 25 de abril de 2011;
 - Págico topics
 - Evaluated answers
 - ..
- Special Edition of the Linguamática journal of April 2012 (Santos et al., 2012) – <http://www.linguamatica.com>

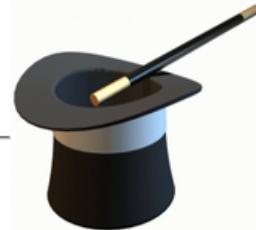


Os apontadores na janela do lado direito não actualizam o lado esquerdo do ecrã.

Sugerimos, pois, que naveguem exclusivamente através dos apontadores do lado esquerdo.

Cartola: Pacote de recursos do Págico

[Págico](#), [Linguateca](#)



O Cartola, que foi disponibilizado no dia 17 de abril de 2012, inclui todos os recursos criados no âmbito do Págico, nomeadamente:

- [coleção do Págico](#), de 681.058 documentos da wikipedia portuguesa de 25 de abril de 2011;
- coleção de tópicos do Págico ([xml](#), [txt](#))
- [monte das respostas avaliadas](#)
- [subcoleção do monte do Págico](#)

Além disso, e para facilitar estudos de aspectos particulares, também disponibilizamos

- [lista de respostas corretas e justificadas, sem as respetivas justificações](#)
- [lista de respostas corretas e justificadas, com as respetivas justificações](#)
- [lista de respostas consideradas corretas independentemente de estarem bem justificadas](#)

E a nova versão do SIGA encontra-se disponível em formato tar.gz: [SIGA_2012.tar.gz](#)

Financiamento

A Linguateca e o Págico foram financiados pelas seguintes entidades até 31 de Dezembro de 2011.



MCTES MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR



Mas continuaram a ser apoiados pelas seguintes instituições:





Pagico_001 pt/1/9/7/1972_filme_.965afb.xml
Pagico_001 pt/a/r/a/Araguaya_-_Conspiração_do_Silêncio.a41d5e.xml
Pagico_001 pt/a/s/_/As_Meninas_filme_.7acb87.xml
Pagico_001 pt/b/a/t/Batismo_de_Sangue.26611b.xml
Pagico_001 pt/b/e/y/Beyond_Citizen_Kane.d35325.xml
Pagico_001 pt/b/r/a/Brazil,_a_Report_on_Torture.b7dae1.xml
Pagico_001 pt/b/r/a/Brazil_Cinema,_Sex_and_the_Generals.f2bdd0.xml
Pagico_001 pt/c/a/b/Cabra-cega_filme_.e29b13.xml
Pagico_001 pt/c/a/b/Cabra_Marcado_para_Morrer.f5574e.xml
Pagico_001 pt/c/o/n/Condor_filme_.c34b5d.xml
Pagico_001 pt/c/o/r/Corpo_em_Delito.072df4.xml
Pagico_001 pt/d/e/d/Dedão_Mamata.214f2f.xml
Pagico_001 pt/f/e/l/Feliz_Ano_Velho_filme_.9f7035.xml
Pagico_001 pt/j/a/n/Jango_filme_.2b9fe4.xml
Pagico_001 pt/l/a/m/Lamarca.259536.xml
Pagico_001 pt/l/u/l/Lula,_o_Filho_do_Brasil_filme_.b218b5.xml
Pagico_001 pt/l/u/t/Lutas_filme_.01bc03.xml
Pagico_001 pt/m/a/r/Marighella_-_Retrato_Falado_do_Guerrilheiro.c697c3.xml
Pagico_001 pt/m/e/m/Memórias_do_Cércere_filme_.d248b4.xml
Pagico_001 pt/o/_/a/O_Ano_em_que_Meus_Pais_Saram_de_Férias.b58539.xml
Pagico_001 pt/o/_/b/O_Bom_Burguês.4c5c73.xml
Pagico_001 pt/o/_/q/O_Que_%_Isso,_Companheiro_filme_.ec72f9.xml
Pagico_001 pt/o/s/_/Os_Herdeiros.a0f25f.xml
Pagico_001 pt/p/i/n/Pindorama_filme_.58777d.xml
Pagico_001 pt/p/r/a/Pra_frente,_Brasil.634849.xml
Pagico_001 pt/q/u/a/Quase_Dois_Irmãos.c80d70.xml
Pagico_001 pt/s/o/n/Sonhos_e_Desejos.e40399.xml
Pagico_001 pt/t/e/r/Terra_em_Transe.d67729.xml
Pagico_001 pt/z/u/z/Zuzu_Angel_filme_.b6838a.xml
Pagico_002 pt/a/_/i/A_Indomada.de74a7.xml
Pagico_002 pt/a/_/v/A_Viagem_1975_.ea152d.xml
Pagico_002 pt/a/_/v/A_Viagem_1994_.3b816d.xml
Pagico_002 pt/c/a/m/Caminhos_do_Coração.c40597.xml
Pagico_002 pt/f/e/r/Fera_Ferida.c92cc4.xml
Pagico_002 pt/k/u/b/Kubanacan.9ff6cd.xml
Pagico_002 pt/o/_/b/O_Beijo_do_Vampiro.5bddf8.xml
Pagico_002 pt/o/_/b/O_Bem-Amado_telenovela_.41a30d.xml
Pagico_002 pt/o/_/c/O_Clone.12457d.xml
Pagico_002 pt/o/_/f/O_Fim_do_Mundo.5d4e8e.xml
Pagico_002 pt/o/s/_/Os_Mutantes_Caminhos_do_Coração.045ff2.xml

Pagico_001: FILMES SOBRE A DITADURA OU SOBRE O GOLPE MILITAR NO BRASIL

Pagico_002: TELENOVELAS BRASILEIRAS DO GÊNERO REALISMO FANTÁSTICO



Pagico_001 pt/1/9/7/1972_filme_.965afb.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/a/r/a/Araguaya_-ConspiraÃ§Ã£o_do_SilÃ¡ncio.a41d5e.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_La
Pagico_001 pt/a/s/_As_Meninas_filme_.7acb87.xml
Pagico_001 pt/b/a/t/Batismo_de_Sangue.26611b.xml
Pagico_001 pt/b/a/t/Batismo_de_Sangue.26611b.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/b/e/y/Beyond_Citizen_Kane.d35325.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/b/r/a/Brazil,_a_Report_on_Torture.b7dae1.xml
Pagico_001 pt/b/r/a/Brazil_Cinema,_Sex_and_the_Generals.f2bdd0.xml
Pagico_001 pt/c/a/b/Cabra-cega_filme_.e29b13.xml
Pagico_001 pt/c/a/b/Cabra-cega_filme_.e29b13.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/c/a/b/Cabra_Marcado_para_Morrer.f5574e.xml
Pagico_001 pt/c/a/b/Cabra_Marcado_para_Morrer.f5574e.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/c/o/n/Condor_filme_.c34b5d.xml
Pagico_001 pt/c/o/r/Corpo_em_Delito.072df4.xml
Pagico_001 pt/d/e/d/DedÃ©_Mamata.214f2f.xml
Pagico_001 pt/f/e/l/Feliz_Ano_Velho_filme_.9f7035.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml pt
Pagico_001 pt/j/a/n/Jango_filme_.2b9fe4.xml
Pagico_001 pt/j/a/n/Jango_filme_.2b9fe4.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml pt/c/i/n/Cir
Pagico_001 pt/l/a/m/Lamarca.259536.xml
Pagico_001 pt/l/a/m/Lamarca.259536.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/l/u/l/Lula,_o_Filho_do_Brasil_filme_.b218b5.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.60551
Pagico_001 pt/l/u/t/Lutas_filme_.01bc03.xml
Pagico_001 pt/m/a/r/Marighella_-_Retrato_Falado_do_Guerrilheiro.c697c3.xml
Pagico_001 pt/m/e/m/MemÃ³rias_do_CÃ¡rcere_filme_.d248b4.xml
Pagico_001 pt/o/_a/O_Ano_em_que_Meus_Pais_SaÃram_de_FÃ©rias.b58539.xml
Pagico_001 pt/o/_a/O_Ano_em_que_Meus_Pais_SaÃram_de_FÃ©rias.b58539.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Lat
Pagico_001 pt/o/_b/O_Bom_BurguÃ¡s.4c5c73.xml
Pagico_001 pt/o/_q/O_Que Ã‰_Isso,_Companheiro_filme_.ec72f9.xml
Pagico_001 pt/o/_q/O_Que Ã‰_Isso,_Companheiro_filme_.ec72f9.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.(
Pagico_001 pt/o/s/_Os_Herdeiros.a0f25f.xml
Pagico_001 pt/o/s/_Os_Herdeiros.a0f25f.xml pt/c/i/n/Cinema_do_Brasil.a12cc6.xml
Pagico_001 pt/p/i/n/Pindorama_filme_.58777d.xml
Pagico_001 pt/p/r/a/Pra_frente,_Brasil.634849.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/q/u/a/Quase_Dois_IrmÃƒos.c80d70.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/s/o/n/Sonhos_e_Desejos.e40399.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/t/e/r/Terra_em_Transe.d67729.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml
Pagico_001 pt/z/u/z/Zuzu_Angel_filme_.b6838a.xml
Pagico_001 pt/z/u/z/Zuzu_Angel_filme_.b6838a.xml pt/a/n/e/Anexo_Lista_de_filmes_sobre_ditaduras_militares_na_AmÃ©rica_Latina.605516.xml



Págico and LogOnto..

- Págico
 - Can it be useful to reasoning systems?
 - Bootstrap a KB from Págico?
 - Test reasoning systems?
- More about Pagico:
 - <http://www.linguateca.pt/Pagico/>

Page type	# Documents
Templates	32 900
Disambiguation	5 006
Redirection	574 077
Media	9 678
Articles	856 005

Table 6: Document distribution per page type.



References

- Diana Santos, Cristina Mota, Cláudia Freitas e Luís Costa (eds.)
Linguamática 4 (1). Abril, 2012
- Acknowledgments

Financiamento

A Linguateca e o Págico foram financiados pelas seguintes entidades até 31 de Dezembro de 2011.



UMIC
Agência para
a Sociedade
do Conhecimento



Fundaçao para a Computação Científica Nacional



MCTES MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR



Fundaçao para a Ciéncia e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Mas continuaram a ser apoiados pelas seguintes instituições:

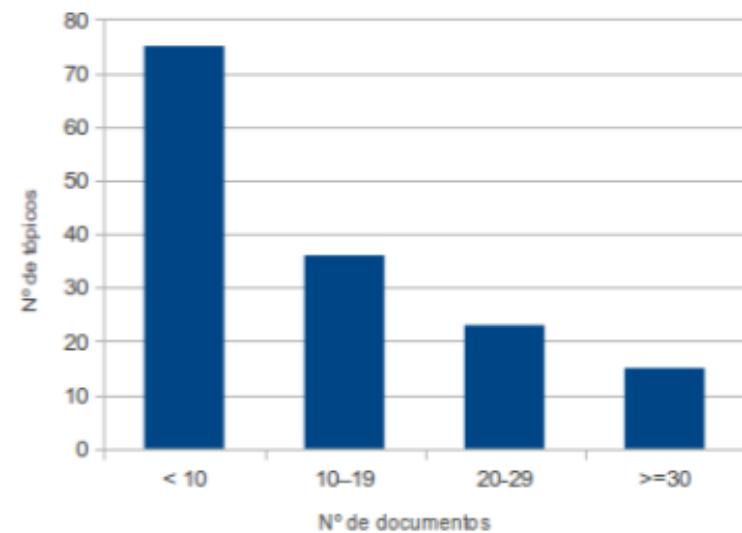




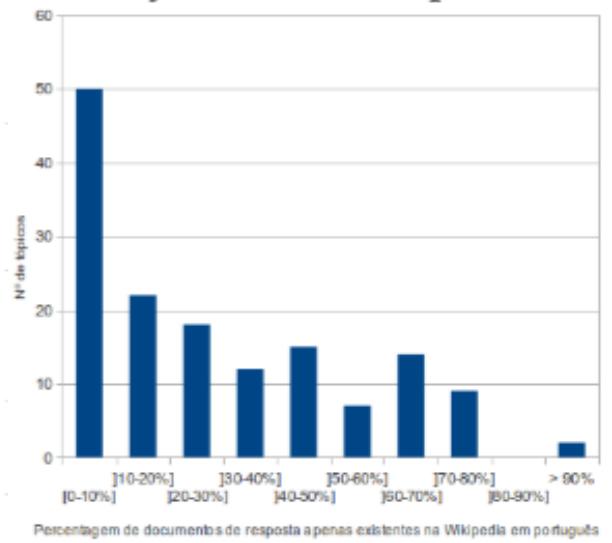
CARTOLA <http://www.linguateca.pt/Cartola>

Págico answers pool

Number of answers and justification documents



Percentage of answers and justifications
only in the PT wikipedia



Págico – Português Mágico



MOTIVATION AND TASK

Is it possible to develop better systems to answer realistic user needs, searching for answers to a particular topic in Wikipedia? Is Wikipedia in Portuguese good enough to provide information on lusophone topics? Can we learn from watching people trying to answer them? Is competition or cooperation between human and automatic participants worth indulging in?



Pagico

Tópicos

[Filmes sobre a ditadura ou sobre o golpe militar no Brasil](#)

Temas

Artes, Letras

Dados geográficos I

Brasil

Pagico_001

Tema(s): Artes, Letras

Dados geográficos: Brasil

[Página principal](#) | [Sair](#) [observador_hum]



Filmes sobre a ditadura ou sobre o golpe militar no Brasil

Tema:

Resposta	Justificação	Correta?	Justificada?
Sonhos e Desejos	Sem Justificações	Sim	Corrigir avaliação
Lamarca	Sem Justificações	Sim	Corrigir avaliação
Quase Dois Irmãos	Sem Justificações	Sim	Corrigir avaliação
As Meninas filme	Sem Justificações	Sim	Corrigir avaliação
Cabra Marcado para Morrer	Sem Justificações	Sim	Corrigir avaliação
Brazil, a Report on Torture	Sem Justificações	Sim	Corrigir avaliação
Lamarca	Sem Justificações	Sim	Corrigir avaliação
Dedé Mamata	Sem Justificações	Sim	Corrigir avaliação
O Ano em que Meus Pais Saíram de Férias	Sem Justificações	Sim	Corrigir avaliação
Cabra-cega filme	Sem Justificações	Sim	Corrigir avaliação
O Bom Burguês	Sem Justificações	Sim	Corrigir avaliação
Marighella - Retrato Falado do Guerrilheiro	Sem Justificações	Sim	Corrigir avaliação
Os Herdeiros	Sem Justificações	Sim	Corrigir avaliação
Zuzu Angel filme	Sem Justificações	Sim	Corrigir avaliação
O Que É Isso, Companheiro filme	Sem Justificações	Sim	Corrigir avaliação
Jango filme	Sem Justificações	Sim	Corrigir avaliação
Batismo de Sangue	Sem Justificações	Sim	Corrigir avaliação



PÁGICO COLLECTION

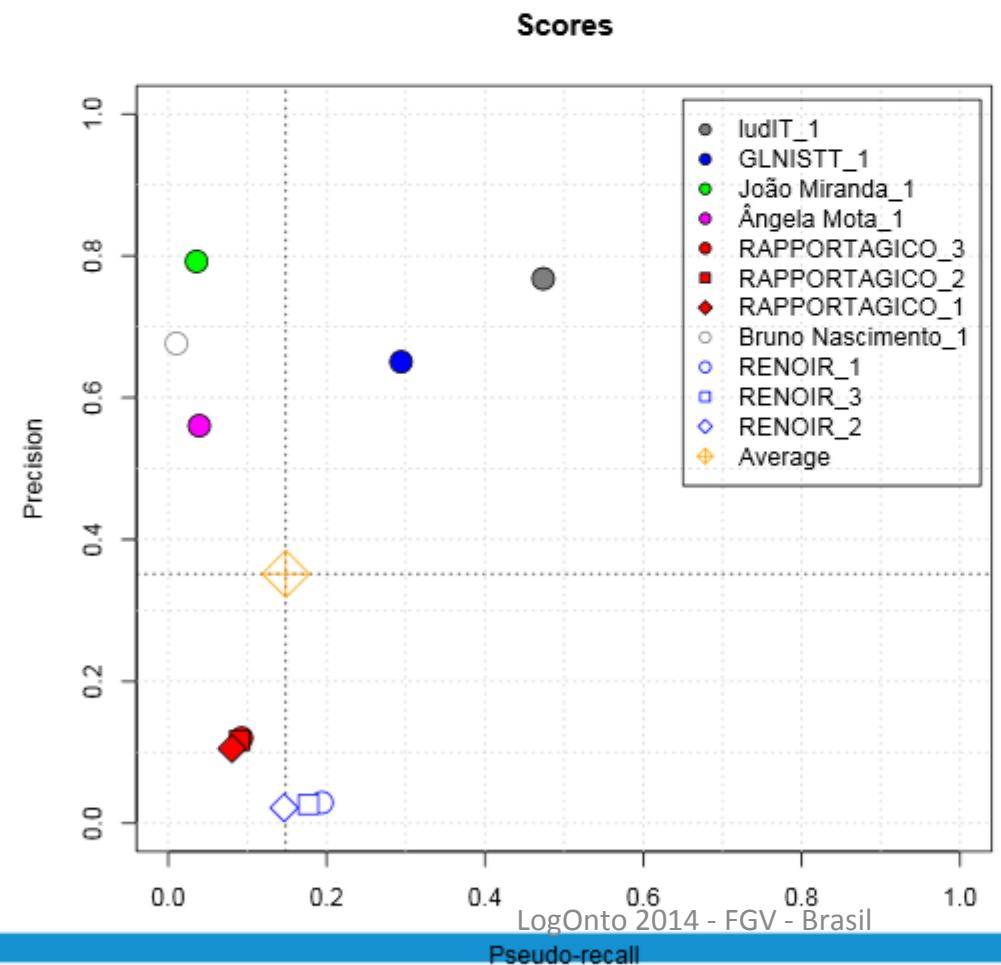
- based on the 25 April 2011 wikipedia snapshot;
- converted to XHTML using:
 - mwlib for the markup conversion;
 - MediaWiki::DumpFile to control the snapshot parsing;
 - in-house tools to manage macro expansion;

Collection constitution:

Page type	Total docs
Template pages	32 900
Disambiguation pages	5 006
Redirection pages	574 077
Multimedia pages	9 678
Article pages	856 005

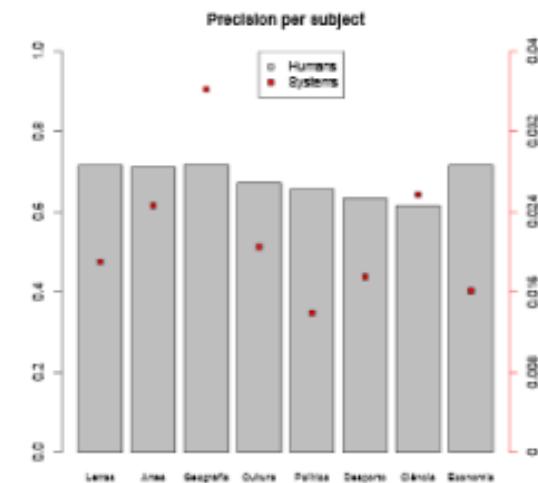
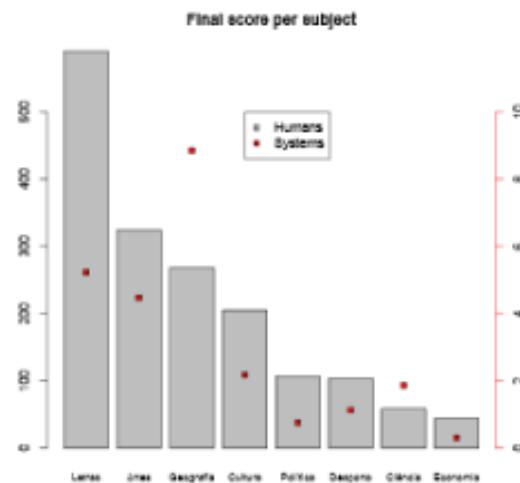
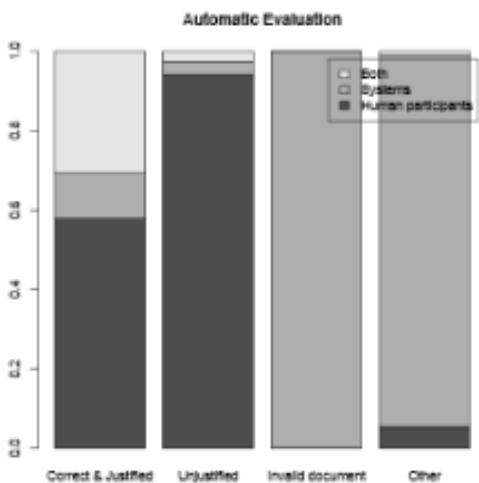
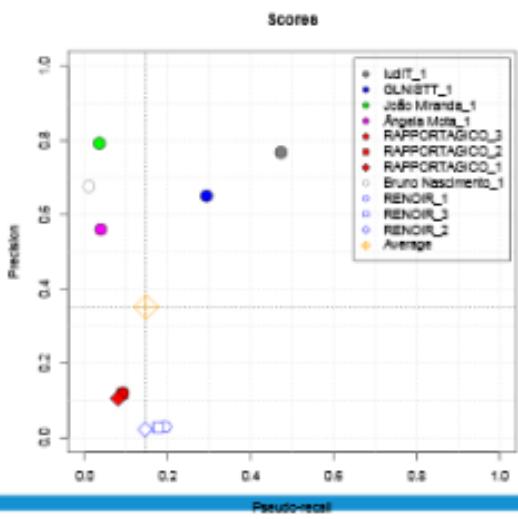


HUMANS VS. SYSTEMS



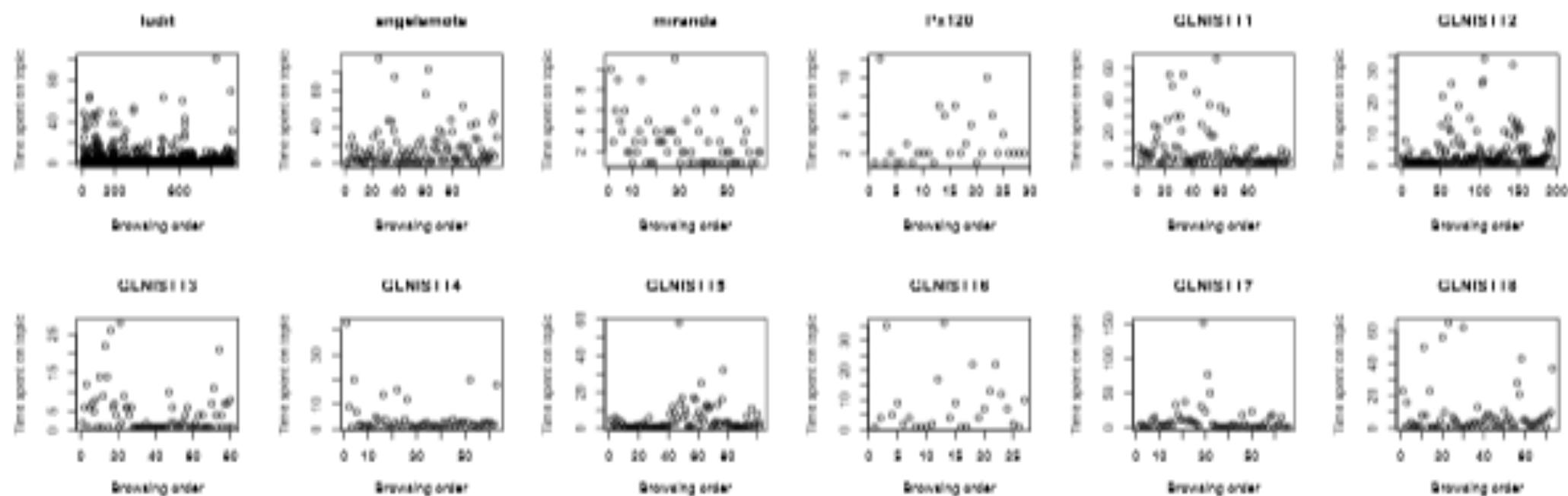


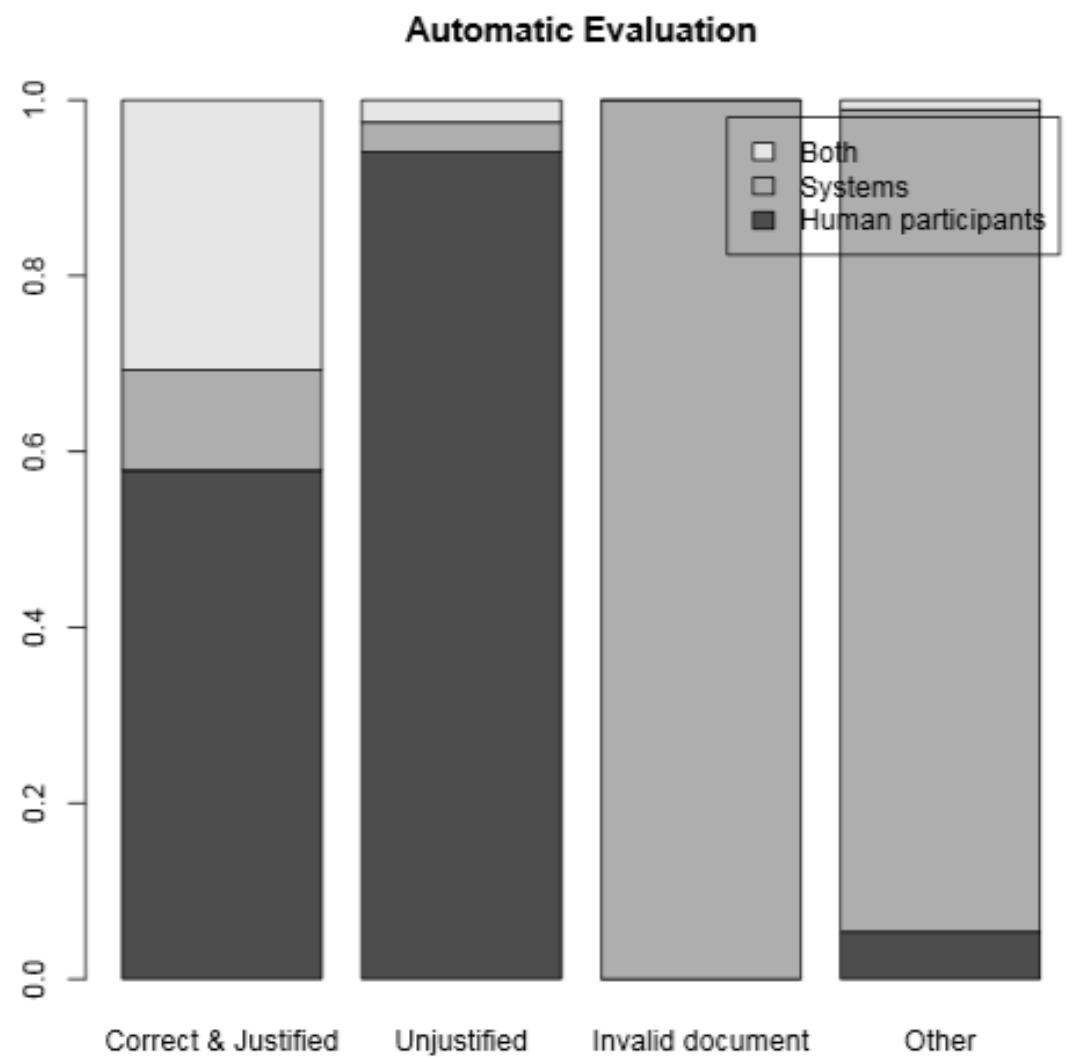
HUMANS VS. SYSTEMS





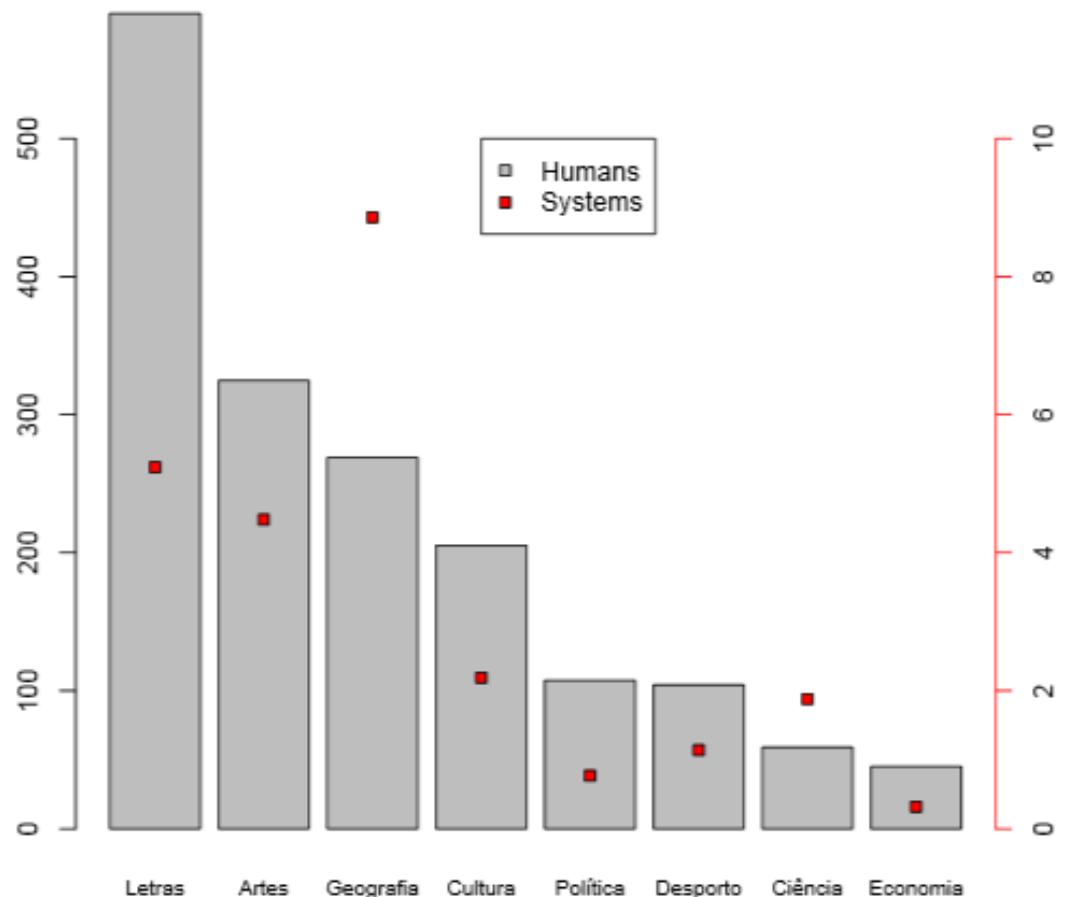
USER BROWSER BEHAVIOUR



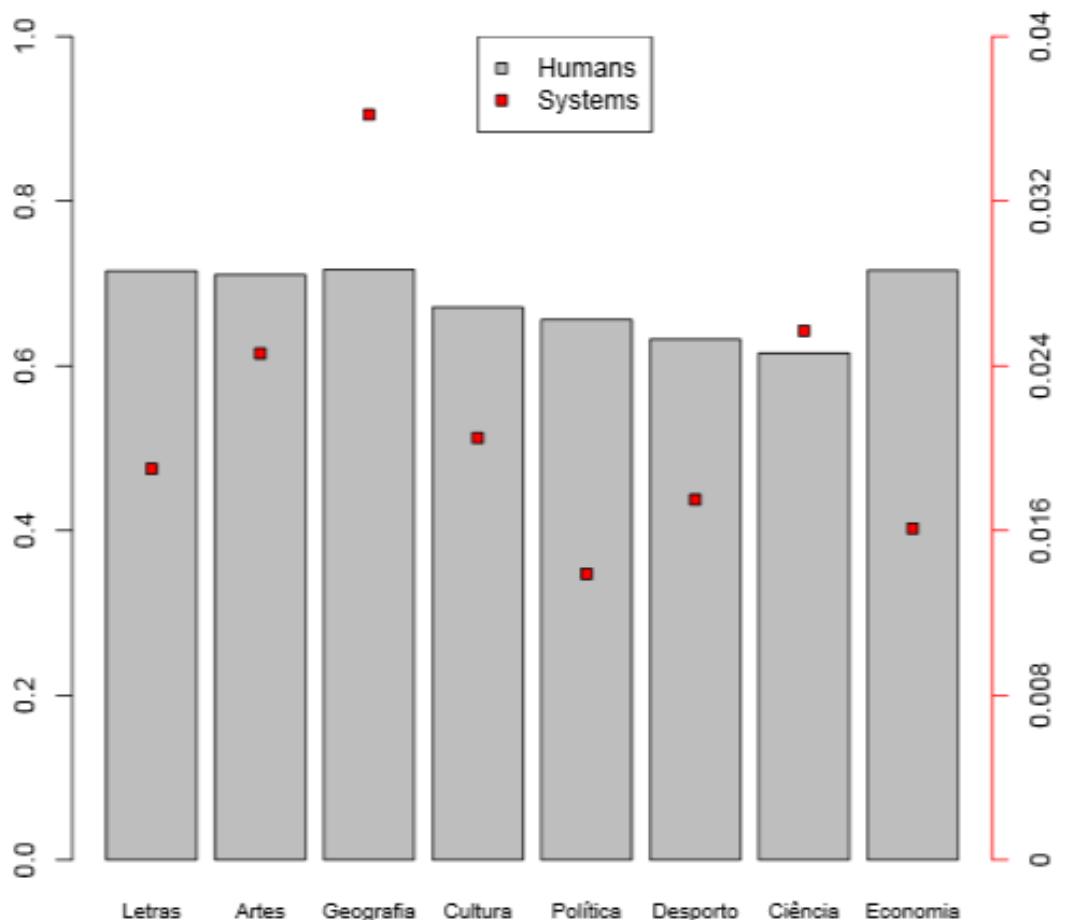




Final score per subject



Precision per subject





EVAL MEASURES

Precision: $P_{p,c} = \frac{|C_{p,c}|}{|R_{p,c}|}$

Pseudo-recall:

$$\alpha_{p,c} = \frac{|C_{p,c}|}{|C_{Pagico}| + |C_{aval}|}$$

Pseudo-F-measure:

$$\phi_{p,c} = 2 \times \frac{P_{p,c} \times \alpha_{p,c}}{P_{p,c} + \alpha_{p,c}}$$

Originality:

$$O_{p,c} = \sum_i^T \sum_j^{R_{p,c,i}} o(r_{p,c,i,j})$$

Creativity:

$$K_{p,c} = \sum_i^T \sum_j^{R_{p,c,i}} k(r_{p,c,i,j})$$

Final score: $M_{p,j} = |C_{p,c}| \times P_{c,j}$

In addition to the measures used in GikiP and GikiCLEF, we chose to investigate originality and creativity, by weighing differently answers according to the number of participants who found them.

SIGA

- Topic creation
- System run submission and testing
- Human participation interface
- Assessment interface
- Conflict resolution
- Pool browsing
- Scoring

Página_001

Filmes sobre a ditadura ou sobre o golpe militar no Brasil

Os filmes podem ser como pano de fundo o período da ditadura.

Documento visualizado: pt/01/Dois_Córregos_film_cae8b6

Dois Córregos (filme)

Dois Córregos é um filme de 1999, dirigido por [Carlos Reichenbach](#) e ambientado no município brasileiro de [Dois Córregos](#).

Sinopse

No inicio do filme, [Beth Goulart](#) tenta de sua adolescência quando vai vir sua propriedade em [Dois Córregos](#). Reichenbach une três mulheres diferentes que convivem com um homem que está escondido por perseguição no período da ditadura. As três mulheres são as atrizes [Inga Liberto](#), [Yanessa Goulart](#) e [Luciana Brandt](#), que também é pianista e interpreta canções na trilha sonora do filme, encabeçada pelo músico e compositor [Ivan Lins](#). O homem é [Carlos Alberto Rascio](#).

Premiações

- Festival Luso Brasileiro: Sônia Maria de Faria: Melhor Atriz, [Inga Liberto](#).
- Festival de Náutol de 1999: Melhor Filme (Jury Oficial), Melhor Tríptico Sonoro ([Ivan Lins](#)), Melhor Atriz Convidada para Luciana Brasil e Melhor Fotografia.
- VII Festival de Cinema e Vídeo de Caieiras: Melhor Filme (Jury Popular), Melhor Diretor e Melhor Atriz para Inga Liberto.
- Prêmio SESC - Os Melhores Do Ano: Melhor Filme (Jury Popular) e Melhor Diretor (Prêmio dos Críticos).

Categoria:Filmes de Carlos Reichenbach Categoria:Filmes de 1999 Categoria:Filmes do Brasil Categoria:Filmes de drama Categoria:Filmes em língua portuguesa

Navegação:
Página_004

ID do tópico: Página_004
Dono do tópico: Dianatuar
Tópico: *Mulheres violenciamos de língua portuguesa*
Descrição do tópico: Clarificação e exemplo de uso
Respostas: [Novas comentários](#) | [Apagar](#)

ID do tópico: Página_004
Dono do tópico: Dianatuar
Tópico: *Mulheres violenciamos de língua portuguesa*
Descrição do tópico: Clarificação e exemplo de uso
Respostas: [Novas comentários](#) | [Apagar](#)

ID do tópico: Página_004
Dono do tópico: Dianatuar
Tópico: *Mulheres violenciamos de língua portuguesa*
Descrição do tópico: Clarificação e exemplo de uso
Respostas: [Novas comentários](#) | [Apagar](#)

ID do tópico: Página_004
Dono do tópico: Dianatuar
Tópico: *Mulheres violenciamos de língua portuguesa*
Descrição do tópico: Clarificação e exemplo de uso
Respostas: [Novas comentários](#) | [Apagar](#)

ID do tópico: Página_004
Dono do tópico: Dianatuar
Tópico: *Mulheres violenciamos de língua portuguesa*
Descrição do tópico: Clarificação e exemplo de uso
Respostas: [Novas comentários](#) | [Apagar](#)

Adicionar ou remover respostas

Documento	Tamanho	Avaliador	Justificativas	(F) Auto-justificado	Remover
pt/01/Comer_Moranga_194196.xml	5,95 KB	Dianatuar	Justificativas (0)	<input type="button" value="Avançar"/>	<input type="button" value="Remover"/>
politic/Denise_Error_6270a.xml	22,64 KB	Dianatuar	Justificativas (0)	<input type="button" value="Avançar"/>	<input type="button" value="Remover"/>
pt/gvt/Golpethmra_Sugia_1a7652.xls	29,18 KB	Dianatuar	Justificativas (0)	<input type="button" value="Avançar"/>	<input type="button" value="Remover"/>

Para encontrar um documento na coleção insira:
 - um URL da Wikipédia (por exemplo, "http://pt.wikipedia.org/wiki/FC_Porto")
 - ou um título da Wikipédia (por exemplo, "[Pedro Nunes \(matemático\)](http://pt.wikipedia.org/wiki/Pedro_Nunes_(matem%C3%A1tico))")
 - ou uma lista de termos que fazem parte do título (por exemplo, [Pedro Nunes matemático](#)).

Página_001
Tópico(s): Artes, Letras
Dados geográficos: Brasil

[Página principal](#) | [Sair](#) [logout]



Filmes sobre a ditadura ou sobre o golpe militar no Brasil

Tópico: Qualquer tópico

[Topos Americanos](#) | [Adicione Tópicos](#)

[Lista de Tópicos](#)

Ir para página visualizada anteriormente

Voltar às respostas a este tópico

Sonhos e Desejos

Sonhos e desejos é um filme brasileiro de 2006, do gênero drama, dirigido por [Marcelo Santiago](#). O roteiro é baseado no romance [Baita do Céu](#), de [Alvam Caldas](#). O filme foi produzido por [Luiz Carlos Barreto](#) Produções Cinematográficas e distribuído por [Caramuru Filmes](#) do Brasil.

Sinopse

Três militantes são obrigados a ficar confinados dentro de um apartamento nos anos 70, em Belo Horizonte. Um deles, um batalhão, acaba de chegar ferido e é assistido e ressuscitado por um caporão. Ele é recebido pela estudante Crisálida e seu professor de literatura Sávio, com quem a jovem tem um romance. Durante a convivência, eles discutem e conversam sobre suas vidas, a política e a realidade da época. Com o tempo, a garota se envolve e se apaixona por Vlady, o [golpista-soldado](#), causando sérios problemas com o narrador.

[Enviar / editar conta](#)

[Artigo](#) [Discussão](#) [Ler](#) [Editar](#) [Ver Histórico](#) [Pesquisa](#)

WIKIPÉDIA A encyclopédia livre

[Página principal](#) [Conteúdo destacado](#)

Sonhos e Desejos

Origem: Wikipedia, a encyclopédia livre.

Sonhos e desejos é um filme brasileiro de 2006, do gênero drama, dirigido por [Marcelo Santiago](#). O roteiro é baseado no romance [Baita do Céu](#), de [Alvam Caldas](#). O filme foi produzido por [Luiz Carlos Barreto](#) Produções Cinematográficas e distribuído por [Caramuru Filmes](#).

[Sonhos e Desejos](#)



Answer evaluation process

- 5 “judges”
- Judgment : **correct /incorrect / doubtful**

- The answer seemed correct, but it wasn't 100% justified
- The page classification wasn't ideal



resposta CORRETA

apropriada

útil

Perspectivas
diferentes

Resposta CORRETA e não-apropriada

Onde fica o Taj Mahal?

- Na 5a Avenida...

(Voorhees e Tice, 2000)



Resposta CORRETA e não-útil

Quem é o autor de Ivanhoe?

- O autor de Rob Roy...

(Spark Jones, 2003)



São respostas “corretas”?

Depende ...

NÃO HÁ UMA RESPOSTA ÚNICA CORRETA



...mas nem tudo está perdido!

as perguntas/tópicos podem ser decompostas em pedaços

Dinossauros carnívoros que habitaram o Brasil

Resposta: X

- X é dinossauro ✓
- X é carnívoro ✓
- X habitou o Brasil ✓

Se algum dos pedaços não está na página resposta, então deve estar em alguma outra página → **JUSTIFICATIVA**

... e as respostas podem ser avaliadas em pedaços

Filmes sobre as relações entre Portugal e suas colônias

Resposta: X

- X é filme ✓
- X é sobre as relações entre Portugal e suas colônias

Devemos exigir?

Como definir?

- Tema principal?
- Romance?
- Comédia?

Os pedaços também podem ser decompostos
em outros pedaços... onde parar?

LogOnto 2014 - FGV - Brasil

Brasil é colônia
Angola é colônia
...



Topic creation

- PT Wikipedia in 150 topics

PT WIKIPEDIA IN 150 TOPICS

Information needs related to Portuguese-speaking countries and their history, with enough coverage in Wikipedia, and not easily browsable through simple categories or infoboxes, spanning areas from History (50) through Geography (26) and Music (19) to Mathematics (1) and Geology (2). How to assess the answers and their justifications was often quite difficult.

Tarefa da organização

- Criar um sistema (SIGA) e preparar a coleção
- Identificar tipos de necessidades de informação com resposta na coleção
- Escolher as respostas
- Avaliar as enviadas pelos participantes
- Definir critérios de avaliação
- Definir medidas
- Criar e disponibilizar recursos