

Resultados págicos: participação, medidas e pontuação

Cristina Mota

Linguateca/FCCN

Encontro do Págico
17 de Abril de 2012, Coimbra

Inscritos

	Tipo de participação	
	Automática	Humana
Inscritos	6	15
- Participantes	2	5
- Desistentes		
responderam a tópicos	-	1
consultaram páginas	-	3
fizeram pesquisas sem consultar	-	2
viram apenas exemplos	2	1
não viram coleções	2	3

Muito entusiasmo, pouca participação

Participantes humanos

Participante	# Respostas	# Com justificação
ludIT	1387	489 (35%)
GLNISTT	1016	255 (25%)
Ângela Mota	157	8 (5%)
João Miranda	101	60 (50%)
Bruno Nascimento	34	1 (3%)
Total	2695	
Distintas	2383	
Total/Distintas	1.13	

ludIT 6 investigadores da área da fala;
Responderam a conjuntos disjuntos de tópicos com colaboração;
Pesquisa na Wikipédia atual e confirmação na coleção do Págico.

GLNISTT 23 estudantes de mestrado divididos em 8 grupos;
Responderam a conjuntos disjuntos de tópicos sem colaboração;
Usaram diferentes fontes, incluindo Wikipédia atual;
Familiarização com avaliações conjuntas e com as tarefas de RI e RAP.

João Miranda Participação individual;
Pesquisa com termos do tópico ou pelo nome do artigo que contém a resposta.

João Miranda; Ângela Mota Participações individuais;
Não enviaram relatos de participação.

Participantes automáticos

Participante (Corrida)	# Respostas
RENOIR (1)	15000
RENOIR (2)	15000
RENOIR (3)	15000
Total	45000
Distintas	28626
Total/Distintas	1.57
RAPPORTAGICO (1)	1718
RAPPORTAGICO (2)	1736
RAPPORTAGICO (3)	1730
Total	5184
Distintas	2343
Total/Distintas	2.21
Total	50184
Distintas	30543
Total/Distintas	1.64

RENOIR Sistema de recuperação de informação geográfica; Documentos mais relevantes para tópicos:

- originais
- lematizados
- lematizados e reformulados

RAPPORTAGICO Reconhecimento de sintagmas fráscicos + identificação de sinónimos usando ontologia lexical.

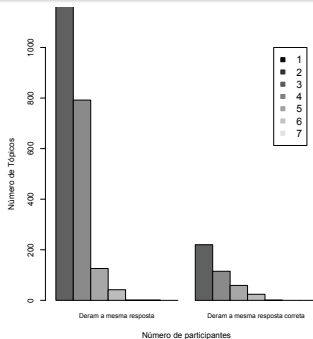
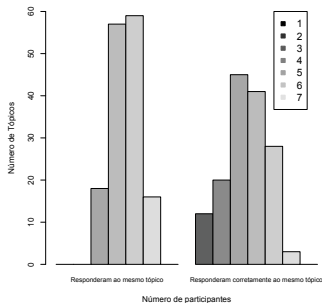
Três corridas:

- Base de comparação,
- Expansão de sinónimos dos sintagmas verbais com:

- *Bag of Words*
- *Personalized Page Rank*.

Resumindo...

	# Respostas
Total	52879
Distintas	32485
Total/Distintas	1.62



- Nenhum tópico foi respondido por todos os participantes
⇒ Nenhuma resposta foi dada por todos os participantes
- Todos os tópicos foram respondidos por pelo menos 3 participantes
- Pouco mais de 10 tópicos foram respondidos corretamente por 1 único participante
- Mais de metade das respostas corretas foram dadas apenas por 1 único participante
- Existe um tópico que não foi respondido corretamente por nenhum participante

Procedimento de avaliação

- 1 Geração do monte
- 2 Avaliação automática
- 3 Distribuição pelos avaliadores
 - de respostas avaliadas automaticamente como corretas mas não justificadas
 - de respostas avaliadas como incorretas
- 4 Avaliação humana
- 5 Cálculo das medidas de avaliação para cada cenário

Medidas de avaliação

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

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

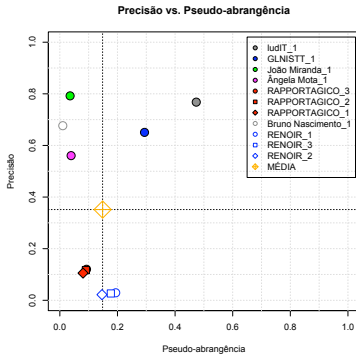
Pseudo-medida-F $\phi_{p,c} = 2 \times \frac{P_{p,c} \times \alpha_{p,c}}{P_{p,c} + \alpha_{p,c}}$

Originalidade $O_{p,c} = \sum_i^T \sum_j^{R_{p,c,i}} o(r_{p,c,i,j})$
⇒ Cada resposta correta original vale o número de participantes que tentou responder

Criatividade $K_{p,c} = \sum_i^T \sum_j^{R_{p,c,i}} k(r_{p,c,i,j})$
⇒ Cada resposta correta vale o número de participantes que tentou responder ponderado pelo inverso do número de participantes que deram a mesma resposta

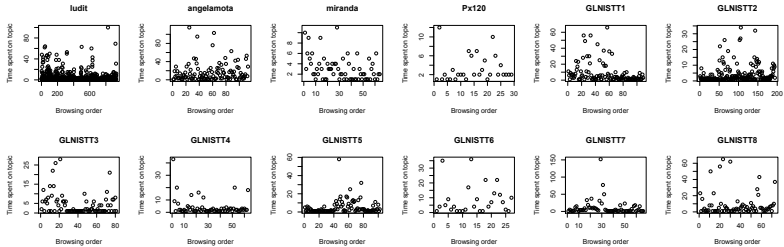
Pontuação final $M_{p,j} = |C_{p,c}| \times P_{c,j}$

Pontuação dos participantes



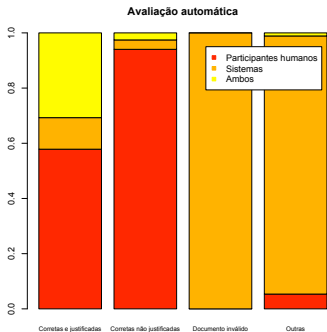
- Melhor precisão para participantes humanos
- Melhor (pseudo-)abrangência também para os participantes humanos!

Tempo e ordem de navegação



- O tempo dispendido em cada tópico não decresceu à medida que se foram lendo mais tópicos
- Será que se perdeu mais tempo nos mesmos tópicos?

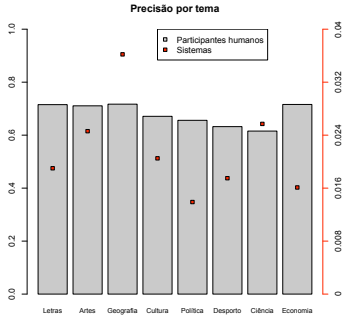
Avaliação automática



- A maior parte das respostas corretas foram dadas por participantes humanos ou por ambos
- Existem 10% de corretas dadas apenas por sistemas
- Quase todas as respostas não justificadas foram dadas apenas por participantes humanos
- A grande fatia por avaliar são respostas de sistemas

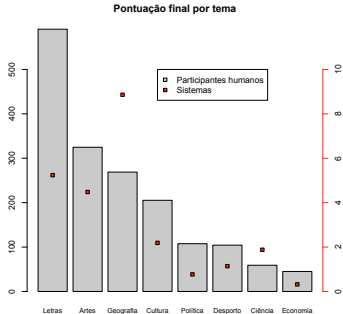
- 288 das 708 respostas dos criadores de tópicos não foram usadas
- ou 184 das 708 respostas, ignorando as justificações ⇒ 104 respostas são iguais mas foram justificadas de modo diferente

Precisão por tema



- Participantes humanos e sistemas são mais precisos em temas diferentes:
 - Participantes humanos mais precisos em Letras, Artes, Geografia e Economia
 - Sistemas mais precisos em Geografia, Ciências e Artes

Pontuação final por tema



- Participantes humanos e sistemas são melhores em temas diferentes:
 - Letras para participantes humanos
 - Geografia para sistemas ⇐ RENOIR apoia-se num sistema de RIG
 - Temas com precisão equivalente têm pontuações finais muito diferentes
 - Letras tem a melhor pontuação final para participantes humanos
 - Economia tem a pior pontuação final para participantes humanos

Comentários finais

- Primeiros passos na comparação entre sistemas e participantes humanos na pesquisa de informação na Wikipédia
- Muito ficou por descobrir, mas o Cartola está disponível para todos poderem responder a perguntas como:
 - Em que é que se distinguem os tópicos com mais e menos respostas para cada tipo de participação?
 - Por que ordem foram percorridos os tópicos relativamente à ordem pré-estabelecida?
 - Que tópicos, temas ou zonas geográficas despertaram mais interesse aos participantes humanos?
 - As medidas de avaliação avaliam bem os participantes?

Linguateca has throughout the years been jointly funded by the Portuguese Government, the European Union (FEDER and FSE), UMIC, FCCN and FCT. Páigico was also supported by the Universities of Oslo, PUC-Rio, Coimbra and FCT grant SFRH/BPD/73011/2010.

