

# Medindo o precipício semântico

Nuno Cardoso

Departamento de Informática

Faculdade de Ciências da Universidade de Lisboa

`ncardoso@xldb.di.fc.ul.pt`

## Resumo

---

Este artigo descreve a minha participação na avaliação conjunta Págico e detalha a estratégia seguida para a participação, que usou um sistema de recuperação de informação geográfica com módulos especializados na interpretação e reformulação semântica de consultas. A estratégia seguida revelou-se demasiado complexa para gerar saídas automáticas, e a participação resumiu-se no envio de três saídas básicas. O artigo faz um resumo das lições aprendidas e tece algumas considerações sobre trabalho futuro.

## Palavras chave

---

Págico, avaliação conjunta, recuperação de informação geográfica, reformulação semântica de consultas

## 1 Introdução

---

A minha participação no Págico foi feita com um sistema de recuperação de informação geográfica (RIG) desenvolvido no âmbito do projecto GREASE (Silva et al., 2006) e da Linguateca (Santos et al., 2004). O projecto GREASE investigou estratégias de adição de raciocínio geográfico em sistemas de recuperação de informação, para que estes obtenham um melhor desempenho na recuperação de documentos para consultas com âmbito geográfico.

Este protótipo RIG participou em várias tarefas de avaliação do GeoCLEF (Gey et al., 2007), com o intuito de avaliar estratégias de combinação de relevância geográfica e relevância textual em consultas típicas de recuperação de informação (Cardoso et al., 2008; Cardoso e Santos, 2008). Contudo, as suas capacidades semânticas ainda eram limitadas, resumindo-se à captura simples de entidades geográficas nas consultas e nos documentos.

Já no âmbito do meu trabalho de doutoramento, o RENOIR foi criado e desenvolvido para participar em avaliações conjuntas piloto como o GikiP (Santos et al., 2009), GikiCLEF (Santos et al., 2010) e NTCIR (Cardoso e Silva, 2010a), onde há um maior foco na interpretação das consultas

e no raciocínio de respostas.

O RENOIR é um módulo de reformulação de consultas que procura compreender e enriquecer as consultas com informação semântica, com o objectivo de tornar a intenção do utilizador mais clara, e melhorando a capacidade do sistema de RIG em recuperar documentos relevantes.

Avaliações como o Págico são de grande relevância para a avaliação de sistemas de RIG, uma vez que é necessário que os sistemas RIG tenham uma capacidade robusta de interpretação de intenções dos utilizadores (como é o caso da deteção de âmbitos geográficos), e porque existe uma forte presença de desafios geográficos na tarefa (Cardoso, 2008a; Santos, Cardoso e Cabral, 2010).

O meus objectivos na participação no Págico são dois: i) medir o desempenho do RENOIR na interpretação dos tópicos e na geração de respostas correctas, e ii) medir o desempenho do sistema de RIG na recuperação de documentos que correspondam a entidades que são respostas correctas.

No entanto, as saídas oficiais enviadas foram obtidas apenas com o sistema de RIG a funcionar sem o auxílio do RENOIR, e em configurações básicas. Apesar da estratégia proposta implicar grandes dificuldades na geração de resultados de forma automática, acredito que um sistema que se proponha realizar uma tarefa como a apresentada pelo Págico, terá de seguir uma estratégia semelhante à que segui, e que passo a explicar de seguida.

## 2 Estratégia semântica proposta

---

O tipo de tópicos usado no Págico e em avaliações similares, embora sejam avaliações que procurem fundir a recuperação de informação (RI) com a resposta a perguntas (RAP), são na sua grande maioria tópicos de RAP, ou seja, perguntas que requerem como resposta uma ou mais entidades (cidades, pessoas, organizações, etc.), coadjuvados se possível com a justificação

para a escolha da(s) resposta(s).

Em avaliações conjuntas para sistemas de RI, os tópicos procuram imitar as necessidades de informação típicas dos utilizadores (Peters e Braschler, 2001; Rachel Aires et al., 2003; Voorhees e Harman, 2005). Normalmente, os tópicos referem um determinado tema de interesse, e espera-se que o sistema de RI encontre documentos relevantes para esse mesmo tema. No caso de avaliações de sistemas RIG, os tópicos incluem um determinado âmbito geográfico. Na prática, os temas escolhidos para os tópicos são fortemente condicionados pelas colecções usadas, uma vez que é preciso existir uma quantidade mínima de documentos relevantes para que o tópico possa ser considerado útil para a avaliação.

Assim sendo, todo o sistema que se proponha participar em avaliações como o Páxico precisa de ter uma noção precisa do que são entidades, como encontrar essas entidades e atribuir classificações semânticas, compreender o papel dessas entidades no contexto da pergunta, e ter a capacidade de calcular a probabilidade de que um outro conjunto de entidades são respostas adequadas para a pergunta inicial.

Por outras palavras, um sistema de RI não tem essa capacidade; Um sistema de RIG também não tem, apesar da sua necessidade de detectar entidades geográficas e de raciocinar sobre o domínio geográfico. Como tal, a primeira tarefa é dotar o sistema de RIG da capacidade de analisar as colecções e consultas a uma maior profundidade semântica. Esse papel é desempenhado pelo RENOIR, que se encontra esquematizado na Figura 1.

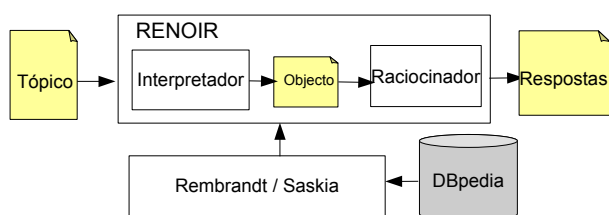


Figura 1: Esquema de funcionamento do reformulador de consultas RENOIR.

O funcionamento do RENOIR é resumido da seguinte forma: o tópico é analisado pelo *interpretador*, que converte as perguntas em *objectos* computacionais, que procuram representar as diversas propriedades da pergunta de uma forma simples e inteligível para que o *raciocinador* possa usar. O *raciocinador* decide qual a melhor estratégia para obter respostas correctas e, se possível, gera essa lista de entidades candidatas a respostas finais.

Para a obtenção de respostas, o raciocinador recorre frequentemente à DBpédia, uma base de dados gerada a partir de instantâneos da Wikipédia (Auer et al., 2007). A DBpédia pode ser acedida em <http://dbpedia.org>, e permite a consulta dos seus dados usando consultas SPARQL<sup>1</sup>. Em resumo, a função do raciocinador é a de traduzir a pergunta inicial, formulada em língua natural, num conjunto de perguntas SPARQL, e verificar se as respostas obtidas correspondem ao tipo de entidades esperadas pela pergunta.

Um exemplo de funcionamento do RENOIR pode ser dado pelo tópico 81, "Igrejas em Macau", que o raciocinador do RENOIR pode converter na seguinte consulta SPARQL:

```

SELECT ?Churches WHERE {
  ?Churches skos:subject
  <http://dbpedia.org/resource/Category:
    Churches_in_Macau>
}
  
```

Na DBpédia 3.7, a consulta gera os seguintes 6 resultados:

```

dbpedia.org/resource/Ji_Dou_Church
dbpedia.org/resource/St._Dominic%27s_Church_%28Macao%29
dbpedia.org/resource/St._Joseph%27s_Seminary_and_Church
dbpedia.org/resource/Ruins_of_St._Paul%27s
dbpedia.org/resource/Macau_Protestant_Chapel
dbpedia.org/resource/St._Lazarus%27_Church
  
```

Vamos analisar ao detalhe os componentes do RENOIR.

## 2.1 Interpretador

O interpretador converte a pergunta, em linguagem natural, num objecto que represente essa pergunta de uma forma facilmente manipulável pelo programa. No processo, o interpretador procura identificar entidades mencionadas, expressões de pergunta (quantos, quais, etc) e outros padrões que possam ser mapeados a entidades na DBpédia.

O objecto é composto pelos seguintes elementos:

**Tema**, a entidade que define o tipo de resposta esperada. O tema pode ser mapeado como i) um recurso da DBpédia com uma propriedade `rdf:type` para um valor `skos:Concept`, ii) uma classe ontológica da DBpédia, ou iii) uma classificação semântica definida pelo HAREM (Santos e Cardoso, 2007; Santos et al., 2008), nesta ordem preferencial.

<sup>1</sup><http://www.w3.org/TR/rdf-sparql-query/>

**Condições**, ou uma lista de critérios que filtram a lista de respostas candidatas. Cada condição é composta por i) uma propriedade ontológica da DBpédia, ii) um operador e iii) um recurso da DBpédia.

**Tipo de Resposta Esperada (TRE)**, que define as propriedades que a lista final de respostas tem que ter.

O objecto é gerado mediante a aplicação de um conjunto de padrões sobre o tópico, previamente anotado pelo analisador morfossintático Palavras (Bick, 2000). Exemplificando com outro tópico do págio, #68: “Bandas brasileiras de punk formadas até 1980”:

**Mapear o tema:** O primeiro conjunto de padrões detecta os termos que definem o tema. No tópico exemplo, a regra “(Que)? <[nome]+ [adjectivo]\*>” captura os termos “bandas brasileiras”, que foram previamente anotados como nome e adjectivos (a presença do termo “Que” é facultativo nesta regra). De seguida, estes termos são mapeados para o recurso da DBpédia [http://dbpedia.org/resource/Category:Brazilian\\_bands](http://dbpedia.org/resource/Category:Brazilian_bands), um recurso criado a partir da respectiva página da Wikipédia para essa categoria (e que, como tal, possui a propriedade `rdf:type` com o valor `skos:Concept`).

**Mapear o TRE:** Depois do tema mapeado, outros padrões determinam o TRE a partir do tipo de pergunta, e o tipo de tema. Para perguntas “Que X” tal como no tópico exemplo, o TRE é atribuído ao tema, ou seja, as respostas têm que ter uma propriedade `skos:subject` com o valor igual ao tema, o que significa que a resposta esperada tem de ser forçosamente uma banda brasileira. Note que, por exemplo, o padrão “Quantos X” atribui o TRE a um número, e diz ao raciocinador que a resposta final tem de ser o tamanho da lista de respostas, ou um valor obtido a partir de uma propriedade DBpédia.

Se o tema não for mapeado a um recurso DBpédia, é então mapeado a uma classe na ontologia DBpédia ou a uma categoria do HAREM, que também pode ser mapeado ao TRE. Supondo que o interpretador não consegue mapear “bandas brasileiras” a um recurso da DBpédia; usando um almanaque interno, o termo “Bandas” faz com que a TRE seja mapeada à classe <http://dbpedia.org/ontology/Band>. Por último, se o interpretador não consegue mapear a uma classe da ontologia DBpédia, é usada a categoria/tipo do HAREM PESSOA/GRUPOIND.

**Mapear restrições:** no tópico exemplo, há duas condições: a primeira filtra as respostas correctas a partir de uma lista inicial de bandas a aquelas que se formaram no Rio de Janeiro, e a segunda filtra as respostas para bandas formadas antes de 1980. O interpretador do RENOIR deve ter um padrão que captura a expressão “formadas? [em|desde|até] X em Y”, que gera duas condições: i) condição formada pela propriedade `dbpedia-owl:yearsActive`, operador `BEFORE`, e um valor `1980-01-01` (data), e ii) condição formada pela propriedade `dbpedia-owl:hometown`, operador `IS`, e uma entidade referente [http://dbpedia.org/resource/São\\_Paulo](http://dbpedia.org/resource/São_Paulo).

## 2.2 Raciocinador

Dependo das propriedades encontradas no objecto, o raciocinador decide qual é a melhor estratégia para obter as respostas. A ação do raciocinador é uma lista de consultas SPARQL realizadas à DBpédia, para obter as respostas e justificações.

No exemplo dado, o objecto ideal teria uma TRE mapeada ao recurso DBpédia que melhore descreve o tipo de entidades que queremos obter como respostas, tal como `Category:Brazilian_punk_rock_groups`. As condições seriam a restrição do local de formação a São Paulo, e a data de formação anterior a 1980. O resultado ideal do raciocinador seria a seguinte consulta SPARQL:

```
SELECT ?x WHERE {
  ?x dct:subject <http://dbpedia.org/resource/Category:Brazilian_punk_rock_groups> .
  ?x dbpedia-owl:hometown
  <http://dbpedia.org/resource/São_Paulo> .
  ?x dbpprop:yearsActive ?y .
  FILTER (?y < "1980-01-01"^^xsd:date) .
}
```

Com a DBpédia 3.7, esta consulta devolve 0 resultados. Relaxando a restrição do ano da formação, a DBpédia devolve dois resultados, [dbpedia.org/resource/Titãs](http://dbpedia.org/resource/Titãs) e [dbpedia.org/resource/Kleiderman](http://dbpedia.org/resource/Kleiderman). Retirando a restrição do local de formação, a DBpédia devolve 10 resultados, o que significa que só 10 bandas são classificadas como bandas de punk brasileiras.

## 3 Integração do RENOIR no sistema de RIG

Como foi exemplificado na secção anterior, a ação do RENOIR pode resultar frequentemente na

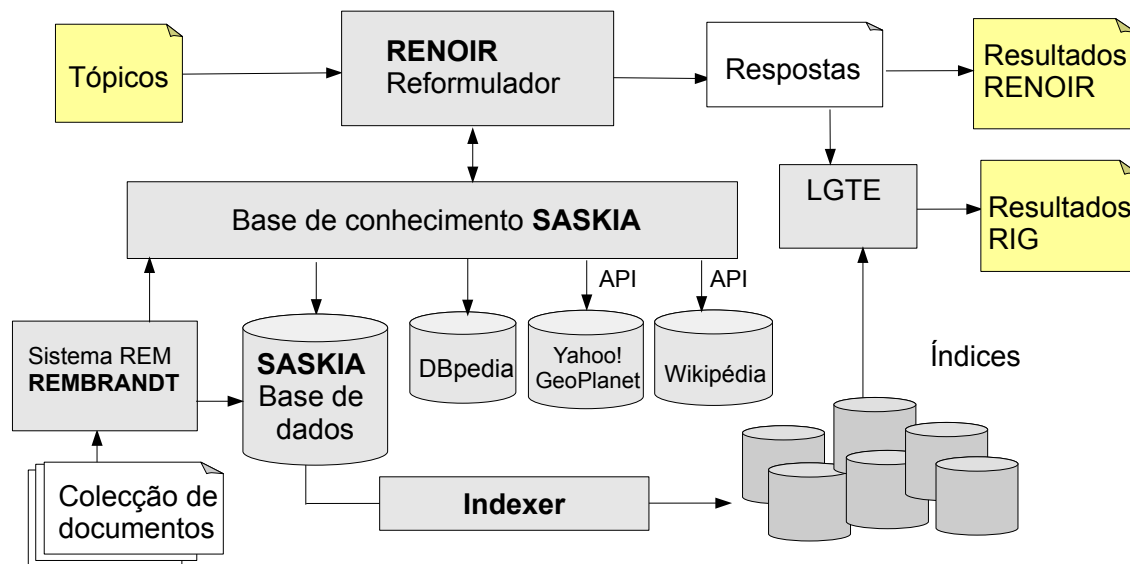


Figura 2: sistema de RIG usado para gerar respostas no Páxico.

geração de nenhuma resposta. Existem muitas razões para tal:

1. O RENOIR não possui padrões suficientes para que a interpretação das perguntas seja razoavelmente abrangente. Este é o principal desafio do RENOIR, que precisa de ser mais tolerante a diferentes formulações de um tipo de conceito;
2. A dificuldade em gerar objectos que representem adequadamente a pergunta inicial. Um exemplo é o tópico 148, “Primeiras universidades de cada país lusófono”. O conceito de universidades primogénitas é difícil de transcrever pelo RENOIR; adicionalmente, o conceito de país lusófono é um conceito que requer um mecanismo de expansão para o formato interno mais conveniente, que é a lista explícita dos países abrangidos.
3. Mesmo assumindo que o objecto está bem descrito, é difícil decidir pelo conjunto de consultas SPARQL certas para a geração de respostas. No caso anterior do tópico #68, o raciocinador poderia optar por gerar consultas SPARQL que usasse uma categoria mais genérica, como por exemplo `Brazilian_music_groups`, e adicionar a restrição:

```
?x dbpedia-owl:genre
<http://dbpedia.org/resource/Punk_Rock>
```

Contudo, esta estratégia iria obter nenhum resultado.

4. Quando o raciocinador obtém nenhum resultado, torna-se difícil de identificar o

que está a provocar tal resultado. Além dos problemas de formulação das consultas SPARQL, há sempre a incerteza se a consulta é demasiado específica. Por exemplo, a informação do ano de formação de bandas pode ser uma informação rara nas páginas da Wikipédia, o que faz que a restrição temporal usada elimine respostas que possam estar correctas só porque a informação não existe por enquanto na DBpédia.

Por outras palavras, em teoria o RENOIR não precisaria do sistema de RIG para gerar respostas válidas; na prática, o RENOIR é demasiado frágil e muito pouco abrangente para gerar respostas de forma consistente, e é necessário recorrer a um sistema de RI/RIG.

Como tal, a participação no Páxico depende mais da capacidade do sistema de RIG em recuperar documentos, do que da capacidade do RENOIR em raciocinar sobre respostas. No entanto, o sistema de RIG usado pode recuperar documentos usando não só a similaridade textual, mas também similaridade em relação a entidades.

O RENOIR consegue reconhecer TRE e entidades nas consultas, e se a colecção de documentos for previamente anotada e as entidades mencionadas devidamente reconhecidas e indexadas, é possível então adicionar uma camada semântica na geração das saídas.

No caso do tópico #120, “Cervejas consumidas em Angola”, o sistema de RIG consegue medir a similaridade geográfica entre o âmbito da consulta (Angola) e os documentos que referem locais em Angola. Por outras palavras, um

	1000 respostas			100 respostas		
	#1	#2	#3	#1	#2	#3
Respostas	15000	15000	15000	1500	1500	1500
correctas, justificadas	436	329	398	129	94	102
correctas, não justificadas	38	25	29	9	11	9
Precisão	0,0291	0,0219	0,0265	0,0860	0,0627	0,0680
Pseudo-abrangência	0,1939	0,1463	0,1770	0,0574	0,0418	0,0454
Pseudo-medida-F	0,0506	0,0381	0,0461	0,0688	0,0501	0,0544
Precisão tolerante	0,0316	0,0236	0,0285	0,0920	0,0700	0,0740
Originalidade	120	151	54	0	0	0
Criatividade	634	514	517	93	83	68
Pontuação final	12,67	7,2	10,6	11,09	5,9	6,9

Tabela 1: Resultados das corridas com base na avaliação de 1000 respostas por tópico, ou de somente 100 respostas por tópico

documento que refere consumo de cerveja em Luanda, tem uma forte possibilidade de ser recuperado, uma vez que o sistema de RIG consegue determinar que Luanda está dentro do âmbito geográfico da consulta, da forma explicada ao detalhe em (Cardoso e Silva, 2010b).

A Figura 2 apresenta o sistema de RIG usado na geração de resultados para o Págico. O sistema baseia-se no modelo clássico de RI, onde os termos dos documentos (ou apenas os seus lemas) são indexados, e a recuperação de documentos faz-se com base no algoritmo BM25 para calcular a similaridade entre os termos dos documentos e os termos da consulta (Robertson et al., 1992). Na participação do Págico, os parâmetros  $k_1=1.2$  e  $b=0.75$  foram usados no algoritmo.

O REMBRANDT é um sistema de reconhecimento de entidades mencionadas (REM), cujo papel no sistema de RIG é a identificação e classificação de todas as entidades mencionadas (EM) presentes na coleção de documentos, e o seu mapeamento a recursos da DBpédia (Cardoso, 2008b; Cardoso, 2012). O REMBRANDT guarda os documentos anotados e as EM reconhecidas na base de dados SASKIA.

A SASKIA organiza as EM em tabelas relacionais para facilitar a geração de índices por parte do indexador (um por cada tipo de EM). A SASKIA também serve como API para diversas fontes de informação como a DBpédia, servindo diversos componentes do sistema como é o caso do raciocinador do RENOIR. Adicionalmente, a SASKIA também associa e armazena informação geográfica da GeoPlanet (Yahoo!, 2011) às EM classificadas como locais, para a posterior geração de índices geográficos, tal como descrito em (Cardoso e Silva, 2010b).

O módulo LGTE (Lucene with GeoTemporal Extensions) (Machado, 2009) é responsável pela

recuperação de documentos, e usa o algoritmo BM25 para calcular a similaridade não só entre termos, mas também entre entidades.

## 4 Participação

A participação no Págico, tal como planeada inicialmente, revelou-se uma tarefa consideravelmente complexa de colocar em prática. A anotação da coleção pelo REMBRANDT revelou-se uma tarefa morosa, tal como a rectificação dos padrões de deteção do RENOIR para a totalidade dos 150 tópicos.

A participação resumiu-se então à geração de saídas usando o sistema de RIG nas suas configurações mais simples. Estas saídas base serão posteriormente usadas para aferir a diferença de desempenho do sistema, quando a anotação da coleção ficar concluída, e o RENOIR conseguir gerar respostas para a maioria dos tópicos.

Assim sendo, as três corridas enviadas são corridas de base:

**Saída #1**, sem nenhum tipo de reformulação de consulta, e usando um índice de termos radicalizados da coleção, e com conversão de diacríticos.

**Saída #2**, sem nenhum tipo de reformulação de consulta, usando um índice de termos não radicalizados da coleção, e sem conversão de diacríticos.

**Saída #3**, com reformulação de consulta, e usando o índice com termos radicalizados e conversão de diacríticos. A reformulação foi feita por com o algoritmo de retorno de relevância cego (*blind relevance feedback*), usando os 10 primeiros documentos no retorno, e adicionando os 16 termos mais relevantes.

O radicalizador usado foi o Snowball (Agichtein e Gravano, 2000). As saídas foram limitadas a 1000 documentos por tópico, e os resultados estão apresentados na tabela 1.

Os resultados mostram que o desempenho do sistema de RIG foi muito fraco, o que reflete o quão inadequado é a utilização de um sistema de RI nesta tarefa sem o auxílio de qualquer tipo de estratégia semântica na procura das respostas.

Mesmo com valores baixos, pode-se observar que as saídas com os índices de termos lematizados geram melhores resultados, uma vez que os termos são agrupados no seu lema, e a recolha é menos sensível às flexões das palavras.

O algoritmo de retorno de relevância cego gera resultados piores, uma vez que é um algoritmo que depende imenso da qualidade dos documentos retornados; se estes são, na sua grande maioria, documentos irrelevantes para o tópico, então os termos gerados poderão dispersar ainda mais o foco da recolha.

## 5 Ilações da participação

A primeira ilação a retirar desta participação, é a nível técnico: o sistema de RIG usado tem problemas de escalabilidade na anotação. A anotação dos documentos permite gerar um dicionário de EM na coleção e permite usar EMs e as suas classificações semânticas no processo de selecção e recolha documentos. Contudo, a coleção do Págico é composta por mais de 500.000 documentos de tamanho considerável, pelo que a anotação requer grandes recursos computacionais. O REMBRANDT é um sistema focado na qualidade de anotação, não na rapidez.

A segunda ilação prende-se com a fragilidade da estratégia do RENOIR, nomeadamente da sua dependência num conjunto de manuais na interpretação e raciocínio de consultas. O sistema de RAP desenvolvido pela Priberam também adota uma estratégia semelhante, e que também requer um esforço considerável para a afinação dos seus padrões sintáticos, de forma a conseguir as elevadas prestações obtidas nas avaliações conjuntas em que participou (Amaral et al., 2009).

Além da fragilidade na interpretação das perguntas, geração dos objetos e raciocínio das respostas, a DBpédia ainda pode ser considerada um projecto em franco desenvolvimento. Cada versão da DBpédia é gerada com base em diferentes instantâneos da Wikipédia inglesa; apesar de recentemente a DBpédia disponibilizar bases de dados com base nos instantâneos da Wikipédia portuguesa, o módulo SASKIA ainda

não está preparado para o utilizar.

Adicionalmente, a DBpédia introduz também alterações significativas nas suas bases de dados, como por exemplo a criação de novas propriedades, ou a revisão das suas classes ontológicas. Tais revisões tornam as regras de raciocínio do RENOIR rapidamente obsoletas.

Estas dificuldades técnicas são incontornáveis em qualquer sistema que se proponha desempenhar a tarefa do Págico. Tais sistemas precisam de: i) obter diversas informações sobre uma grande quantidade de entidades em frações de segundos, usando um leque de recursos como almanaques, ontologias ou bases de dados de conhecimento como a DBpédia; ii) compreender as perguntas apresentadas em linguagem natural, e usar a informação disponível para raciocinar e obter respostas, tal como um humano. Veja-se o exemplo do sistema de RAP Watson da IBM (Ferrucci, 2011).

Assim sendo, e uma vez que as dificuldades técnicas estarão sempre presentes, resta mitigar esses problemas e focar na avaliação da estratégia semântica, e na sua aplicabilidade num sistema de RIG. Como trabalho futuro, o REMBRANDT será adaptado para anotações mais rápidas e menos completas, para poder lidar com coleções de grande tamanho. Para concluir a avaliação no Págico, um subconjunto de 15 tópicos do Págico (mais relevantes para os objectivos de avaliação de RIG) serão usados para a avaliação pós-Págico.

## Agradecimentos

Agradeço a Diana Santos e a José João de Almeida, pelos comentários e sugestões de melhoria deste artigo. Este trabalho foi suportado pela FCT pelo financiamento anual ao LASIGE, projecto GREASE-II (PTDC/EIA/73614/2006) e bolsa de doutoramento SFRH/BD/45480/2008, governo português, União Europeia (FEDER e FSE) através do projecto Linguateca, segundo o contracto ref. POSC/339/1.3/C/NAC, UMIC e FCCN.

## Referências

- Agichtein, Eugene e Luis Gravano. 2000. Snowball: Extracting Relations from Large Plain-Text Collections. Em *Proceedings of the 5th ACM Conference on Digital Libraries (DL'00)*, pp. 85–94, San Antonio, TX, EUA, June 2-7, 2000. ACM.
- Amaral, Carlos, Adán Cassan, Helena Figueira,

- André Martins, Afonso Mendes, Pedro Mendes, José Pina, e Cláudia Pinto. 2009. Priberam's question answering system in qa@clef 2008. Em *Proceedings of the 9th Cross-language evaluation forum conference on Evaluating systems for multilingual and multimodal information access*, CLEF'08, pp. 337–344, Berlin, Heidelberg. Springer-Verlag.
- Auer, Sören, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, e Zachary Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. (4825):722–735.
- Bick, Eckhard. 2000. *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Tese de doutoramento, University of Aarhus, Aarhus, Dinamarca, November, 2000.
- Cardoso, Nuno. 2008a. Novos rumos para a recuperação de informação geográfica em português. Em Diana Santos, editor, *Linguatca: 10 anos. Encontro satélite do PROPOR 2008*, Aveiro, Portugal, 11 de Setembro, 2008. Linguatca.
- Cardoso, Nuno. 2008b. REMBRANDT - Reconhecimento de entidades mencionadas Baseado em relações e análise detalhada do texto. Em Cristina Mota e Diana Santos, editores, *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*. Linguatca.
- Cardoso, Nuno. 2012. Rembrandt - a named-entity recognition framework. Em *Proceedings of the eighth International Conference on Language Resources and Evaluation, LREC 2012*, Istambul, Turquia, 21–27 de Maio, 2012. a aguardar publicação.
- Cardoso, Nuno, David Cruz, Marcirio Chaves, e Mário J. Silva. 2008. Using Geographic Signatures as Query and Document Scopes in Geographic IR. 5152:802–810.
- Cardoso, Nuno e Diana Santos. 2008. To separate or not to separate: reflections about GIR practice. Em *Proceedings of the 1st Workshop on Novel Methodologies for Evaluation in Information Retrieval, NMEIR'2008*, Glasgow, UK, 30 March, 2008.
- Cardoso, Nuno e Mário J. Silva. 2010a. Experiments with Semantic-flavored Query Reformulation of Geo-Temporal Queries. Em *Working Notes of the 8th NTCIR Workshop*, Tóquio, Japão, 15–18 de Junho, 2010.
- Cardoso, Nuno e Mário J. Silva. 2010b. A GIR Architecture with Semantic-flavored Query Reformulation. Em *6th Workshop of Geographic Information Retrieval, GIR 10*, Zurique, Suíça, 18-19 de Fevereiro, 2010.
- Ferrucci, David A. 2011. Ibm's watson/deepqa. *SIGARCH Computer Architecture News*, 39(3).
- Gey, Fredric, Ray Larson, Mark Sanderson, Kerstin Bishoff, Thomas Mandl, Christa Womser-Hacker, Diana Santos, Paulo Rocha, Giorgio Di Nunzio, e Nicola Ferro. 2007. GeoCLEF 2006: the CLEF 2006 Cross-Language Geographic Information Retrieval Track Overview. 4730:852–876.
- Machado, Jorge. 2009. LGTE: Lucene Extensions for Geo-Temporal Information Retrieval. Em *Workshop on Geographic Information on the Internet Workshop (GIIW), held at ECIR 2009*, Toulouse, França, 9 de Abril, 2009.
- Peters, Carol e Martin Braschler. 2001. Cross-Language System Evaluation: the CLEF campaigns. *Journal of the American Society for Information Science and Technology*, 52(12):1067–1072.
- Rachel Aires, Sandra Aluísio, Paulo Quaresma, Diana Santos, e Mário J. Silva. 2003. An initial proposal for cooperative evaluation on information retrieval in Portuguese. Em Jorge Baptista, Isabel Trancoso, Maria das Graças Volpe Nunes, e Nuno J. Mamede, editores, *Computational Processing of the Portuguese Language: 6th International Workshop, PROPOR 2003*, pp. 227–234, Faro, Portugal, Junho, 2003. Springer Verlag.
- Robertson, Stephen E, Steve Walker, Micheline Hancock-Beaulieu, Aarron Gull, e Marianna Lau. 1992. Okapi at TREC-3. Em *Proceedings of the 3rd Text REtrieval Conference*, pp. 21–30, Gaithersburg, MD, USA.
- Santos, Diana, Luís Miguel Cabral, Corina Forascu, Pamela Forner, Fredric Gey, Katrin Lamm, Thomas Mandl, Petya Osenova, Anselmo Peñas, Álvaro Rodrigo, Julia Schulz, Yvonne Skalban, e Erik Tjong Kim Sang. 2010. GikiCLEF: Crosscultural Issues in Multilingual Information Access. Em Nicoletta et al. Calzolari, editor, *Proceedings of the 7th conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, Maio, 2010. European Language Resources Association (ELRA).
- Santos, Diana e Nuno Cardoso, editores. 2007. *Reconhecimento de entidades mencionadas*

*em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área.* Linguateca.

[geoplanet/guide/concepts.html](http://geoplanet/guide/concepts.html). acessido em Fevereiro de 2012.

Santos, Diana, Nuno Cardoso, e Luís Miguel Cabral. 2010. How geographic was gikiclef?: a gir-critical review. Em Ross Purves, Paul Clough, e Christopher B. Jones, editores, *GIR*. ACM.

Santos, Diana, Nuno Cardoso, Paula Carvalho, Iustin Dornescu, Sven Hartrumpf, Johannes Leveling, e Yvonne Skalban. 2009. GikiP at GeoCLEF 2008: Joining GIR and QA forces for querying Wikipedia. Em Carol Peters, Tomas Deselaers, Nicola Ferro, Julio Gonzalo, Gareth J.F.Jones, Mikko Kurimo, Thomas Mandl, Anselmo Peñas, e Viviane Petras, editores, *Evaluating Systems for Multilingual and Multimodal Information Access: 9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008, Aarhus, Denmark, September 17-19, 2008, Revised Selected Papers*. Springer.

Santos, Diana, Paula Carvalho, Hugo Oliveira, e Cláudia Freitas. 2008. Second HAREM: new challenges and old wisdom. (5190):212–215.

Santos, Diana, Alberto Simões, Ana Frankenberg-Garcia, Ana Pinto, Anabela Barreiro, Belinda Maia, Cristina Mota, Débora Oliveira, Eckhard Bick, Elisabete Ranchhod, José João Dias de Almeida, Luís Cabral, Luís Costa, Luís Sarmento, Marcirio Chaves, Nuno Cardoso, Paulo Rocha, Rachel Aires, Rosário Silva, Rui Vilela, e Susana Afonso. 2004. Linguateca: um centro de recursos distribuído para o processamento computacional da língua portuguesa. Em Guillermo De Ita Luna, Olac Fuentes Chávez, e Mauricio Osorio Galindo, editores, *Proceedings of the international workshop “Taller de Herramientas y Recursos Lingüísticos para el Espanol y el Portugués” and IX Iberoamerican Conference on Artificial Intelligence, IBERAMIA 2004*, pp. 147–154, Puebla, México, Novembro, 2004.

Silva, Mário J., Bruno Martins, Marcirio Chaves, Ana Paula Afonso, e Nuno Cardoso. 2006. Adding Geographic Scopes to Web Resources. *CEUS - Computers, Environment and Urban Systems*, 30:378–399.

Voorhees, Ellen M. e Donna Harman, editores. 2005. *TREC: Experiment and Evaluation in Information Retrieval*. MIT Press.

Yahoo! 2011. GeoPlanet™Key Concepts. <http://developer.yahoo.com/geo/>