

## **Apêndice G**

# **SAHARA - Serviço de Avaliação HAREM Automático**

Nuno Cardoso

**Nota das editoras:** Este apêndice foi preparado especialmente para o presente livro, depois de ter surgido a ideia no Encontro do Segundo HAREM de implementar um serviço de avaliação na rede. Este serviço, criado por Nuno Cardoso, ficou patente ao público desde 4 de Outubro de 2008.

---

O SAHARA, disponível em [www.linguateca.pt/HAREM](http://www.linguateca.pt/HAREM) → Avaliador, é um serviço na rede que permite a avaliação imediata de saídas de sistemas de REM de acordo com o ambiente de avaliação usado no Segundo HAREM. O SAHARA facilita consideravelmente a avaliação de sistemas de REM, visto que elimina a necessidade de executar uma sucessão de comandos específicos de cada programa de avaliação para obter um conjunto de valores de desempenho. O SAHARA permite ainda a comparação imediata com os resultados oficiais do Segundo HAREM, bem como o acesso aos resultados de cada programa de avaliação, para depuração mais detalhada.

Uma avaliação no SAHARA decorre em três passos:

1. Validação da corrida enviada pelo utilizador, de acordo com o formato do Segundo HAREM
2. Configuração da avaliação pretendida, ou seja, a definição dos cenários, modos de avaliação, e colecções a serem utilizadas
3. Apresentação dos resultados, com um conjunto de tabelas e gráficos que resumem o desempenho do sistema.

## G.1 Primeiro passo: validação

O SAHARA começa por pedir o ficheiro criado pelo sistema de REM (no formato específico de XML do Segundo HAREM), que pode encontrar-se no computador do utilizador, ou num dado URL (ver figura G.1). O avaliador aceita entradas comprimidas por ZIP ou GZIP, desde que incluam o ficheiro XML com o mesmo nome do ficheiro enviado. O formato desse ficheiro é então validado, invocando, entre outros testes, o validador do HAREM que recorre a regras de RelaxNG (van der Vlist, 2003). Só se não forem encontrados erros se avança para o passo da configuração. Caso contrário, o avaliador volta ao passo inicial, mostrando adicionalmente uma lista de erros devidamente localizados no XML com a indicação do número da linha e da coluna, para facilitar a depuração.

## G.2 Configuração da avaliação

O formulário de configuração está dividido em quatro secções: i) Selecção das pistas, ii) Escolha dos cenários, iii) Selecção do modo de avaliação e iv) Escolha da colecção dourada, que serão descritas em seguida.

### G.2.1 Selecção das pistas

A pista clássica do HAREM, descrita no capítulo 1, é obrigatória e está incluída em todas as opções da caixa de selecção (ver figura G.2). Contudo, o utilizador pode optar por

## SAHARA

Serviço de Avaliação HAREM Automático / *HAREM's Automatic Evaluation Service*

[HAREM](#), [Linguateca](#)

Este serviço permite avaliar a saída do seu sistema REM de acordo com os recursos de avaliação (coleções douradas) e as directivas do Segundo HAREM, assim como comparar com os participantes oficiais.  
*This service provides the evaluation of the output of your NER system according to the Second HAREM guidelines and evaluation resources, as well as comparison with official participation.*

**Primeiro passo:** Valide a sua participação.

*Ist step: validate your output.*

Ficheiro (local) (local) file:

ou/or...

Ficheiro (remoto) (remote) file:

Tipos de ficheiros aceites / *File types accepted: XML, XML.GZ, .XML.ZIP*  
O nome do ficheiro de saída deve ser igual ao nome do ficheiro comprimido / *The output filename must be the same as the zipped filename.*

Última actualização / Last update: 11/11/2008.

[Contacte a organização do Segundo HAREM](#) / [Contact us](#)

Figura G.1: Página de entrada do SAHARA.

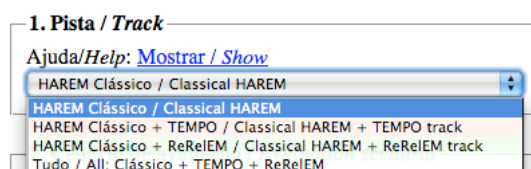


Figura G.2: Configuração da avaliação: selecção das pistas.

incluir também na avaliação as outras pistas: a pista do TEMPO e a pista do ReRelEM, descritas nos capítulos 2 e 4, respectivamente. Para o conjunto de pistas escolhidas, só é possível escolher uma única coleção dourada; para usar diferentes coleções douradas, é preferível realizar uma avaliação de cada vez.

A cada pista está associada uma cor para que seja mais fácil fazer corresponder as várias opções no formulário à pista respectiva.

**2. Cenário de avaliação / Evaluation scenario**  
Ajuda/Help: [Mostrar/Show](#)

**2.1 HAREM clássico / Classical HAREM**

Cenário selectivo da participação / *Participation's scenario*

Cenários do Segundo HAREM / *Second HAREM scenarios*  
 Cenários personalizados / *Custom scenarios*

LOCAL(FISICO{\*};HUMANO{\*});PESSOA(\*)

Categorias / *Categories*: [ABSTRACCAO](#) | [ACONTECIMENTO](#) | [COISA](#) | [LOCAL](#) | [OBRA](#) | [ORGANIZACAO](#) | [PESSOA](#) | [TEMPO](#) | [VALOR](#)

Tipos / *Types*: [Todas / All](#) | [FISICO{\\*}](#) | [HUMANO{\\*}](#) | [VIRTUAL{\\*}](#) | [OUTRO](#)

Cenário selectivo de avaliação / *Evaluation scenario*

Igual ao cenário selectivo do participante / *Same as the participation's scenario*  
 Um novo cenário de avaliação / *New evaluation scenario*

**2.2 Pista ReReIEM / ReReIEM track**

Cenários do ReReIEM do Segundo HAREM / *Second HAREM's ReReIEM scenarios*  
 Cenários personalizados do ReReIEM / *Custom ReReIEM scenarios*

inclui;incluido;ident

Relações / *Relations*: [Inclui/Incluido](#) | [Ident](#) | [Ocorre em/Sede de](#) | [Outra](#)

Figura G.3: Configuração da avaliação: Escolha dos cenários.

### G.2.2 Escolha dos cenários

Os cenários de avaliação permitem escolher o leque de categorias a usar, tanto do lado do participante (Cenário selectivo da participação), como no lado da avaliação (Cenário selectivo de avaliação) – ver figura G.3. Do lado do participante, o formulário permite escolher entre cenários pré-definidos (usados para a geração de resultados oficiais do Segundo HAREM) e cenários personalizados. O utilizador pode especificar o leque de categorias, tipos e subtipos que quer ver avaliados. Do lado da avaliação, é possível escolher entre usar o mesmo cenário especificado no lado do participante, ou então outro cenário personalizado. A avaliação por cenários foi descrita nas secções 5.1.3 e 5.6.3.

Caso se opte por avaliar na pista do ReReIEM, aparece uma caixa para a escolha de cenários específicos para o ReReIEM, descritos na secção 5.8.6. O funcionamento é semelhante à escolha dos cenários para a pista clássica. O utilizador pode escolher cenários para o ReReIEM já usados no Segundo HAREM, ou então criar o seu próprio cenário personalizado.

### G.2.3 Selecção do modo de avaliação

Os modos de avaliação são parâmetros avançados de configuração da avaliação (ver figura G.4). Para a pista clássica, o utilizador pode escolher a forma de tratamento das alternati-

**3. Modo de avaliação / Evaluation mode**

3.1 Pista clássica / classic track

ALT estrita / strict ALT

**Pesos / weights:** Identificação / Identification = 1     $\alpha = 1$      $\beta = 0.5$      $\gamma = 0.25$

3.2 Pista TEMPO / TEMPO track

Estendido completo / Full TEMPO

3.3 Pista ReRelEM / ReRelEM track

ReRelEM com expansão da participação / ReRelEM with run expansion

Figura G.4: Configuração da avaliação: selecção do modo de avaliação.

vas <ALT>, apresentada nas secções 5.1.4 e 5.6.7, entre a forma restrita e a forma relaxada, e também especificar os pesos na fórmula da medida de classificação.

Para a pista do TEMPO, o utilizador pode optar por usar o leque completo de categorização TEMPO (Estendido completo), escolhido por omissão, ou então optar pelo modo sem normalização ou pelo modo só com normalização. A avaliação da pista do TEMPO foi descrita nas secções 5.2 e 5.7.

Para a pista do ReRelEM, cuja avaliação foi apresentada nas secções 5.3 e 5.8, o utilizador pode escolher uma avaliação onde as relações são expandidas automaticamente (escolha por omissão), ou então realizar uma avaliação sem expansão de relações, do lado da corrida do participante.

#### G.2.4 Escolha da colecção dourada

A última secção do formulário de configuração diz respeito à escolha da colecção dourada para a avaliação. As escolhas da colecção estão restritas (por enquanto) às colecções douradas usadas no Segundo HAREM e, consoante as pistas de avaliação escolhidas, ao pedaço da colecção dourada que possui anotações para as pistas seleccionadas.

A colecção dourada do Segundo HAREM possui 129 documentos, onde um subconjunto de 30 documentos foi anotado com todos os atributos da categoria TEMPO e, dentro desse subconjunto, 12 documentos foram anotados com relações para a pista do ReRelEM. Como tal, um utilizador que pretenda avaliar nas pistas clássica + TEMPO, pode escolher entre o subconjunto TEMPO ou o subconjunto ReRelEM, mas não pode escolher a colecção dourada completa.

### G.3 Apresentação dos resultados

De acordo com as pistas seleccionadas e a configuração escolhida, o SAHARA desenha um plano de execução para a avaliação pretendida, e executa os diferentes programas em sequência, que pode ser visualizada (através de uma série de caixas) enquanto o utili-

**4. Colecção dourada / Golden collection**

4.1 HAREM clássico / *Classical HAREM*

CD do Segundo HAREM para ReReLEM / Second HAREM GC for ReReLEM

4.2 Pista TEMPO / *TEMPO track*

CD do Segundo HAREM para ReReLEM / Second HAREM GC for ReReLEM

4.3 Pista ReReLEM / *ReReLEM track*

CD do Segundo HAREM para ReReLEM / Second HAREM GC for ReReLEM

Figura G.5: Configuração da avaliação: escolha da colecção dourada.

**SAHARA**Serviço de Avaliação HAREM Automático / *HAREM's Automatic Evaluation Service*

Versão 1.0

[HAREM, Linguateca](#)

Aguarde por favor. Dependendo do tamanho da corrida, do tipo de avaliação seleccionado e da carga no servidor, a geração de resultados pode demorar vários segundos.  
 Please wait. Generation of the results may take several seconds, depending on the run size, the kind of evaluation selected and the server load.

A executar o expansor ReReLEM...

HAREM clássico / *Classical HAREM*.

CD/GC: CDSegundoHAREM\_ReReLEM.xml.

Modo/Mode: ALT estrito / *Strict ALT*.

Cenário selectivo do participante/Participation's scenario: \*.

Cenário selectivo da avaliação/Evaluation scenario: \*.

Pesos/Weights:  $\text{Id}=1$ ,  $\alpha=1$ ,  $\beta=0.5$ ,  $\gamma=0.25$ .Pista TEMPO / *TEMPO track*.

CD/GC: CDSegundoHAREM\_ReReLEM.xml.

Modo/Mode: Estendido completo / *Full TEMPO*.Pista ReReLEM / *ReReLEM track*.

CD/GC: CDSegundoHAREM\_ReReLEM.xml.

Modo/Mode: ReReLEM com expansão da participação / *ReReLEM with run expansion*.

Cenário/Scenario: \*.



Figura G.6: Execução da avaliação.

zador espera pelos resultados, como ilustrado na figura G.6. Os parâmetros escolhidos na configuração são novamente apresentados, para que se possa rever o tipo de avaliação associada aos resultados a gerar.

Uma vez terminada a geração de resultados, o SAHARA verifica se há resultados oficiais que sejam comparáveis à avaliação realizada, isto é, se existem sistemas participantes no Segundo HAREM que tenham sido avaliados conforme o mesmo cenário de avaliação, o mesmo modo de avaliação e a mesma coleção dourada. (De qualquer maneira, note-se que os resultados oficiais foram gerados com os pesos por omissão, e que a alteração dos pesos desvirtua a comparação.) Se existirem resultados oficiais, o gráfico-resumo dos resultados para cada pista apresenta também os resultados para os três melhores sistemas (considerando apenas a melhor saída de cada sistema), como apresentado na figura G.7.

## SAHARA

Serviço de Avaliação HAREM Automático / HAREM's Automatic Evaluation Service

Versão 1.0

[HAREM, Linguatca](#)

[Enviar uma nova saída / Submit another run](#)

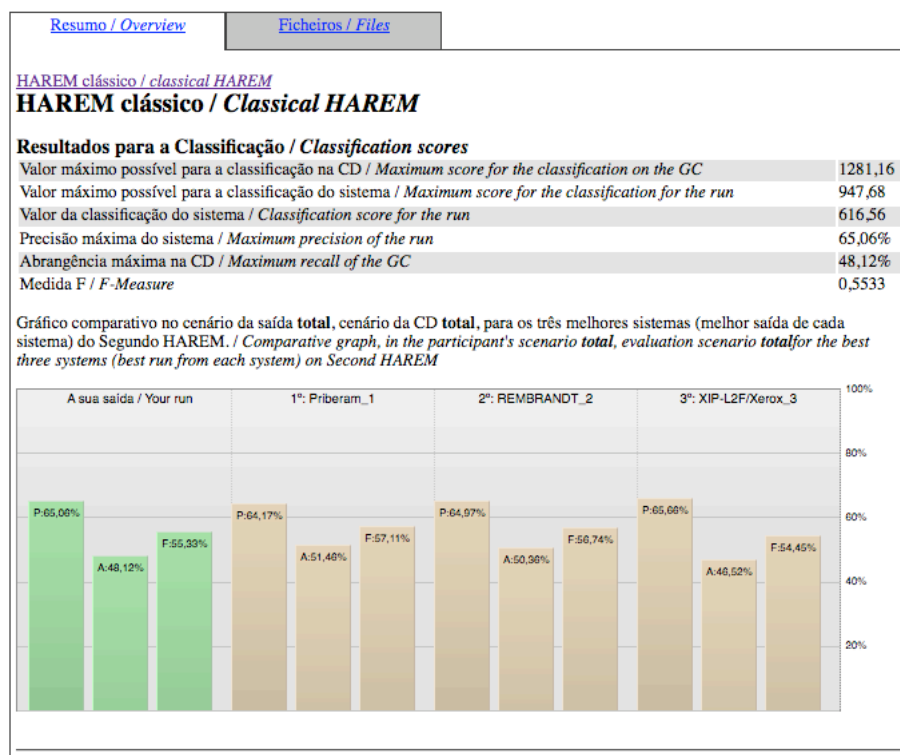


Figura G.7: Apresentação de resultados: tabela de resultados e gráfico comparativo.

Para uma análise mais detalhada dos resultados da avaliação, o utilizador pode escolher o separador 'Ficheiros / Files' para ter acesso à saída de cada programa de avaliação invocado pelo SAHARA (ver figura G.8). Adicionalmente, é possível clicar em 'Depuração dos programas de avaliação' para voltar a ver a sequência de execução dos programas. Ao clicar em cada caixa do diagrama, o SAHARA mostra a linha de comandos executada, o que

é útil para quem pretenda ir buscar os programas de avaliação e executá-los numa linha de comandos.

## SAHARA

Serviço de Avaliação HAREM Automático / *HAREM's Automatic Evaluation Service*

Versão 1.0

[HAREM, Linateca](#)

[Enviar uma nova saída / Submit another run](#)

<a href="#">Resumo / Overview</a>	<a href="#">Ficheiros / Files</a>
<p><b>HAREM clássico / Classical HAREM:</b>  <a href="#">AlinhEM</a> - Alinhador / <i>Aligner</i>.  <a href="#">AvalIDA</a> - Avaliador da identificação / <i>Identification scorer</i>.  <a href="#">Véus</a> - Filtros / <i>Filters</i>.  <a href="#">ALTina</a> - Organizador de ALT / <i>ALT Organizer</i>.  <a href="#">Emir</a> - Avaliador da classificação / <i>Classification scorer</i>.  <a href="#">Ida</a> - Resumidor de classificações / <i>Classification aggregator</i>.</p> <p><b>Pista TEMPO / TEMPO track:</b>  <a href="#">AvalTEMPO</a> - Avaliador de TEMPO / <i>TEMPO scorer</i>.  <a href="#">IdaTEMPO</a> - Resumidor de classificações do TEMPO / <i>TEMPO classification aggregator</i>.</p> <p><b>Pista ReRelEM / ReRelEM track:</b>  <a href="#">Expansor M2</a> - Expansor para o ReRelEM / <i>ReRelEM expander</i>.  <a href="#">Seleccção</a> - Seleccção para o ReRelEM / <i>ReRelEM selector</i>.  <a href="#">Normalizador</a> - Normalizador das ID do ReRelEM / <i>ReRelEM ID normalizer</i>.  <a href="#">Triplas</a> - Conversor de notação para triplas do ReRelEM / <i>ReRelEM format converter to triples</i>.  <a href="#">Filtro</a> - Filtro do ReRelEM / <i>ReRelEM filter</i>.  <a href="#">Avaliador</a> - Avaliador do ReRelEM / <i>ReRelEM scorer</i>.  <a href="#">Resumidor</a> - Resumidor de classificações ReRelEM / <i>ReRelEM classification aggregator</i>.</p>	
<p>Depuração dos programas de avaliação / <i>Evaluation script debug</i>: <a href="#">Mostrar / Show</a>  Mostrar/Show</p>	

Figura G.8: Lista de ficheiros de saída dos programas de avaliação.