

Glossário

Paula Carvalho, Cristina Mota, Diana Santos, Cláudia Freitas
e Hugo Gonçalo Oliveira

Cristina Mota e Diana Santos, editoras, *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*, 2008, p. [v-ix](#).

- Abrangência** Métrica que afere a qualidade da participação, em termos da proporção de respostas correctas no universo de respostas possíveis.
- Análises alternativas** Diferentes análises concorrentes, dentro do mesmo ALT, associadas a uma dada sequência.
- ALT** Etiqueta que identifica um conjunto de análises alternativas associadas a uma mesma sequência, passível de ser segmentada de formas distintas.
- Atributos do HAREM clássico** Conjunto de atributos previstos nas directivas do HAREM clássico, designadamente, CATEG, TIPO e SUBTIPO, os quais também são adoptados nas directivas do TEMPO para caracterizar as expressões temporais.
- Atributos do ReReIEM** Conjunto de atributos especificamente utilizados na pista do ReReIEM, designadamente, TIPOREL e COREL.
- Atributos do TEMPO estendido** Conjunto de atributos específicos da categoria TEMPO, designadamente, TEMPO_REF, SENTIDO, VAL_NORM e VAL_DELTA.
- Avaliação conjunta** Modelo de avaliação que permite a comparação do progresso dos sistemas participantes numa dada área, usando, para tal, um conjunto de recursos comum e métricas e medidas bem definidas.
- Avaliação estrita de ALT** Contabilização de todas as análises alternativas possíveis para um dado segmento do texto, recebendo cada elemento do ALT um peso igual ao inverso do número de alternativas associadas a esse segmento.
- Avaliação relaxada de ALT** Avaliação do HAREM clássico onde, dentro de cada ALT, é seleccionado apenas o elemento que maximiza o valor da medida de classificação do sistema, tal como foi feito no Primeiro HAREM.
- Cenário selectivo** Subconjunto de categorias, tipos, subtipos e/ou outros atributos do cenário total definido pelo participante ou pela organização.
- Cenário de avaliação** Subconjunto de categorias, tipos, subtipos e/ou outros atributos do cenário total sobre o qual vai incidir a avaliação. Pode coincidir com o cenário total.
- Cenário de participação** Subconjunto de categorias, tipos, subtipos e/ou outros atributos do cenário total em que um dado sistema participou e pretende ser avaliado. Pode coincidir com o cenário total.
- Cenário total** Conjunto total de categorias, tipos e subtipos propostos no Segundo HAREM.
- Classificação** Classificação das EM de acordo com as categorias, tipos, subtipos e/ou outros atributos previstos nas directivas.
- Colecção dourada (CD) do Segundo HAREM** Subconjunto de documentos da colecção HAREM anotados de acordo com as directivas do HAREM clássico (e do TEMPO quanto aos atributos CATEG, TIPO e SUBTIPO), e com base nos quais os sistemas participantes são avaliados no HAREM clássico, que inclui a avaliação da pista do TEMPO em relação aos atributos mencionados.

- Colecção dourada do ReReEM (ou CD do ReReEM)** Subconjunto de documentos da CD do Segundo HAREM anotados de acordo com as directivas do ReReEM, e com base nos quais os sistemas participantes são avaliados na tarefa de reconhecimento de relações entre EM.
- Colecção dourada do TEMPO (ou CD do TEMPO)** Subconjunto de documentos da CD do Segundo HAREM anotados de acordo com as directivas do TEMPO, e com base nos quais os sistemas participantes são avaliados na pista do TEMPO.
- Colecção do Segundo HAREM** Conjunto de documentos, com géneros e registos variados, que os sistemas participantes tiveram de anotar no âmbito do Segundo HAREM.
- Corrida no Segundo HAREM (ou corrida)** Colecção do Segundo HAREM anotada por um sistema participante. O mesmo que **participação no Segundo HAREM**.
- Directivas do HAREM clássico** Conjunto de regras e indicações que foram adoptadas na identificação e classificação das EM, no âmbito da pista do HAREM clássico, e segundo as quais os sistemas foram avaliados.
- Directivas do ReReEM** Conjunto de regras e indicações que foram adoptadas na identificação e classificação das relações entre EM, no âmbito da pista do ReReEM, e segundo as quais os sistemas foram avaliados.
- Directivas do TEMPO** Conjunto de regras e indicações que foram adoptadas na identificação, classificação e normalização das expressões temporais, no âmbito da pista do TEMPO, e segundo as quais os sistemas foram avaliados.
- Elemento do ALT** Elemento do conjunto de análises alternativas possíveis.
- Encaixe (de EM)** Existência de EM dentro de uma EM maior.
- Entidade mencionada (EM)** No HAREM clássico, as EM correspondem, em geral, a uma entidade referida no texto por um nome próprio ou por uma expressão numérica, cuja caracterização mais pormenorizada se encontra nas directivas.
- EM espúria** Sequência identificada pelos sistemas participantes como EM, que não se encontra anotada na CD.
- EM vaga** Entidade mencionada que tem mais de uma interpretação (e como tal recebe mais de uma classificação no HAREM). Cada interpretação é designada por faceta.
- Expansão de relações** Explicitação das consequências das propriedades de simetria, existência de inversa e transitividade, e aplicação das regras de expansão, a um conjunto de relações numa colecção.
- Faceta** Interpretação parcial do conteúdo de uma EM vaga, seleccionando apenas uma categoria (ou tipo, ou subtipo) do conjunto de categorias (ou tipos, ou subtipos) postulado para essa EM.
- HAREM** Acrónimo para “HAREM - Avaliação de Reconhecimento de Entidades Mencionadas”, uma série de avaliações conjuntas na área do reconhecimento de entidades mencionadas em português.

HAREM clássico Pista de avaliação de REM no âmbito do Segundo HAREM, baseada nas directivas do HAREM clássico, cujo objectivo é anotar as EM de acordo com o leque de categorias, tipos e subtipos especificados nas directivas. A avaliação do HAREM clássico tem também em conta a avaliação da pista do TEMPO para os atributos do HAREM clássico.

HAREM clássico na CD do TEMPO Uso da CD do TEMPO como base de avaliação segundo as directivas do HAREM clássico e do TEMPO quanto aos atributos do HAREM clássico.

HAREM clássico na CD do ReReIEM Uso da CD do ReReIEM como base de avaliação segundo as directivas do HAREM clássico e do TEMPO quanto aos atributos do HAREM clássico.

Identificação Reconhecimento de que uma dada expressão constitui uma EM, independentemente da sua classificação.

Medida Função que permite a transformação quantitativa da pontuação num valor numérico.

Medida-F Métrica que combina a precisão e a abrangência para cada tarefa, de acordo com a fórmula estabelecida.

Métrica Conjugação dos valores produzidos por uma dada medida num único valor acumulado.

Mini-HAREM Segundo evento de avaliação do Primeiro HAREM.

Modo de avaliação Selecção de atributos do TEMPO que são alvo de avaliação. Existem quatro modos de avaliação: clássico, estendido completo, estendido sem normalização e estendido só com normalização.

Participação no Segundo HAREM (ou participação) Colecção do Segundo HAREM anotada por um sistema participante. O mesmo que **corrida no Segundo HAREM**.

Peso Valor que é atribuído a de cada um dos atributos de uma EM ou a cada um dos elementos do ALT, no cálculo da medida atribuída a uma dada EM.

Pistas do Segundo HAREM Conjunto das diferentes pistas integradas no Segundo HAREM, designadamente, o HAREM clássico, a pista do ReReIEM e a pista do TEMPO.

Pista do ReReIEM Pista específica do Segundo HAREM, que tem por finalidade o reconhecimento de relações entre entidades mencionadas.

Pista do TEMPO Pista específica do Segundo HAREM, baseada num conjunto de directivas independentes das directivas do HAREM clássico, as directivas do TEMPO (Hagège et al., 2008), que tem em vista o reconhecimento e a normalização de expressões temporais.

Pontuação Avaliação qualitativa da relação entre a resposta do sistema e o que está na CD.¹

Precisão Métrica que afere a qualidade da participação, em termos da proporção de respostas correctas dentro do total de respostas dadas.

Primeiro HAREM Primeira edição do HAREM, que englobou duas avaliações distintas: o primeiro evento do Primeiro HAREM e o Mini-HAREM.

Reconhecimento de entidades mencionadas (REM) Tarefa de identificação e classificação automática de entidades mencionadas.

Reconhecimento de relações entre entidades mencionadas (ReReLEM) Tarefa de identificação e classificação de um conjunto de relações entre EM, previamente estabelecidas nas directivas do ReReLEM.

Relação entre EM Ligação entre duas entidades, de acordo com os critérios sugeridos nas directivas do ReReLEM.

ReReLEM Acrónimo para Reconhecimento de Relações entre Entidades Mencionadas.

Saída Resultado de um sistema. No contexto do Segundo HAREM, uma saída refere-se à anotação da colecção do Segundo HAREM, mas distingue-se de corrida por esta última ser uma saída que foi seleccionada pelo participante para ser avaliada no contexto do HAREM.

Sistema participante Sistema que participou no Segundo HAREM.

Segundo HAREM Segunda edição do HAREM.

TEMPO clássico Modo de avaliação da pista do TEMPO que tem em consideração apenas o preenchimento dos atributos CATEG, TIPO e SUBTIPO.

TEMPO estendido completo Modo de avaliação da pista do TEMPO em que todos os atributos de TEMPO são avaliados, incluindo os atributos específicos do TEMPO.

TEMPO estendido sem normalização Modo de avaliação da pista do TEMPO em que todos os atributos de TEMPO são avaliados, excepto os referentes à normalização.

TEMPO estendido só com normalização Modo de avaliação da pista do TEMPO em que são avaliados os atributos de TEMPO previstos no HAREM clássico conjuntamente com os atributos de normalização.

¹ No Segundo HAREM, adoptámos o trio de termos do Primeiro HAREM “pontuação”, “medida” e “métrica”, embora consideremos agora que os termos “apreciação qualitativa”, “pontuação” e “medida”, respectivamente, teriam sido uma melhor escolha.