

Capítulo 5

Avaliação à medida no Segundo HAREM

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

Cristina Mota e Diana Santos, editoras, *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*, 2008, Capítulo 5, p. 97–129.

Um dos desafios que se coloca na realização de uma avaliação conjunta é o de definir uma forma de avaliar cada participação, por um lado, e o de comparar o seu desempenho com o desempenho das demais participações, pelo outro.

Esse desafio é ainda maior se, por um lado, a avaliação incluir várias pistas com objectivos distintos, e, por outro, existirem sistemas muito diferentes entre si, cada um com uma finalidade específica, por vezes até peculiar, como aconteceu no Segundo HAREM. Basta referir que, em dez participantes, não houve sequer dois que executassem exactamente a mesma tarefa.

A característica que mais se evidencia na avaliação do Segundo HAREM é então a possibilidade de poder comparar as várias participações com base em diferentes vistas sobre a mesma colecção dourada (CD), a que chamamos cenários (selectivos). Essa flexibilidade vem aliás do Primeiro HAREM, tendo sido aperfeiçoada na actual avaliação e aplicada às várias pistas do Segundo HAREM. De facto, como veremos, a avaliação do Primeiro HAREM foi o ponto de partida para a avaliação do Segundo HAREM, muito particularmente no que respeita ao HAREM clássico. Naturalmente, as novas pistas obrigaram ao desenvolvimento de novos programas de avaliação, e alguns programas que adoptámos do Primeiro HAREM também sofreram várias alterações, em função das mudanças introduzidas nesta edição.

5.1 Avaliação do HAREM clássico

5.1.1 Pontuações

No Segundo HAREM, cada entidade mencionada (EM) pode receber uma de três pontuações, no que respeita à sua identificação: *Correcta*, *Em falta* e *Espúria*. Estas pontuações são igualmente utilizadas para pontuar a classificação de cada atributo das EM.

5.1.2 Uma única medida

No Primeiro HAREM existiam apenas dois níveis de classificação: categorias e tipos. De acordo com esses níveis, foi definido um conjunto de quatro medidas (plana, só tipos, só categorias e classificação semântica combinada - CSC) para a avaliação do reconhecimento de entidades mencionadas (ver Santos et al. (2007)). Essas medidas eram baseadas no pressuposto de que todas as EM estavam classificadas obrigatoriamente com categoria e tipo.

No Segundo HAREM, o esquema de anotação adoptado define uma hierarquia de quatro níveis (se tivermos em conta somente o HAREM clássico): identificação da entidade (delimitando-a simplesmente com as etiquetas EM), e preenchimento dos respectivos atributos categoria, tipo e subtipo. Além disso, não é obrigatória a marcação dos níveis mais baixos.

Assim, para esta avaliação, foi definida uma única medida, mais robusta e abrangente, que é aplicada a cada entidade correctamente identificada. A medida engloba os quatro níveis da hierarquia, possibilitando a atribuição de diferentes pesos a cada um desses níveis e ainda a penalização por classificações erradas.

Basta identificar correctamente a EM para se receber o valor 1. O valor total da medida é obtido somando esse valor às parcelas relativas aos níveis da classificação. A medida tem em conta várias questões:

$$1 + \sum_{i=1}^N \left(\left(1 - \frac{1}{n_{cats}}\right) \cdot cat_{certoi} \cdot \alpha + \left(1 - \frac{1}{n_{tipos}}\right) \cdot tipo_{certoi} \cdot \beta + \left(1 - \frac{1}{n_{sub}}\right) \cdot sub_{certoi} \cdot \gamma \right) - \sum_{i=0}^M \left(\frac{1}{n_{cats}} \cdot cat_{esp_i} \cdot \alpha + cat_{certoi} \cdot \frac{1}{n_{tipos}} \cdot tipo_{esp_i} \cdot \beta + tipo_{certoi} \cdot \frac{1}{n_{sub}} \cdot sub_{esp_i} \cdot \gamma \right) \quad (5.1)$$

$$K_{certoi} = \begin{cases} 1 & \text{se o atributo } K_i \text{ estiver correcto,} \\ 0 & \text{se } K_i \text{ estiver incorrecto ou omisso} \end{cases}$$

$$K_{esp_i} = \begin{cases} 1 - K_{certoi} & \text{se o atributo } K_i \text{ estiver preenchido} \\ 0 & \text{se } K_i \text{ estiver omisso} \end{cases}$$

$K \in \{cat, tipo, sub\}$

N = número de diferentes classificações vagas na CD, de acordo com o cenário selectivo.

M = número de classificações espúrias na participação, de acordo com o cenário selectivo.

α, β, γ = parâmetros correspondentes aos pesos das categorias, tipos e subtipos.

Figura 5.1: Fórmula de avaliação no Segundo HAREM

- O preenchimento de cada atributo por parte de um sistema é sempre opcional.
- O peso dado a cada atributo é tanto maior quanto mais difícil for acertar o valor do atributo, ou seja, é proporcional ao número de possibilidades que existem para o preencher. Por outro lado, quanto à penalização aplica-se o critério inverso. Isto é, quanto mais difícil for acertar o preenchimento de um atributo, menor é a penalização por ter o atributo espúrio.
- O número de categorias (n_{cats}) pode variar de acordo com o cenário selectivo que se está a avaliar, sendo no cenário total do Segundo HAREM igual a 10^1 e na avaliação de apenas uma categoria igual a 1 (não existindo, neste caso, valorização adicional pela classificação). O número de tipos (n_{tipos}) e de subtipos (n_{sub}) pode variar, respectivamente, de acordo com o valor da categoria ou categoria e tipo a que se referem, por um lado, e ainda com o cenário selectivo escolhido, pelo outro.
- Cada classificação vaga de uma EM é pontuada de forma independente. Isto leva a que a única forma possível de obter o valor máximo seja o sistema ter classificado a EM vaga exactamente da mesma forma que esta se encontra na CD (isto é, sem classificações espúrias e com todas as diferentes classificações possíveis correctas).
- É feita a distinção entre não conseguir atribuir o valor a um atributo (quando esse atributo é omisso) e assumir que o valor desse atributo não está contemplado no conjunto de categorias, tipos e/ou subtipos propostos nas directivas (classificando-o como OUTRO). Considera-se assim que todos os tipos podem ter um SUBTIPO="OUTRO". Desta forma, quando a EM que se está a avaliar pertence a uma categoria e tipo sem subtipo definido, a última parcela toma o valor 0.

¹ Recordar-se o elenco das dez categorias do Segundo HAREM: ABSTRACCAO, ACONTECIMENTO, COISA, LOCAL, OBRA, ORGANIZACAO, PESSOA, TEMPO, VALOR e OUTRO.

- A penalização resultante do preenchimento espúrio de um atributo nunca contém mais de uma parcela. Ou seja, para calcular a penalização é apenas contabilizada a parcela do atributo espúrio no nível mais alto da hierarquia. Por exemplo, se uma classificação tiver a categoria correcta e o tipo espúrio, o subtipo será obrigatoriamente também espúrio. No entanto, a penalização será dada apenas pelo tipo espúrio e não pelo subtipo.

Para a avaliação oficial do Segundo HAREM optou-se por dar maior importância à classificação da categoria do que à classificação do tipo, que por sua vez deveria ter mais importância do que a classificação do subtipo.

Assim, aos atributos categoria, tipo e subtipo foram atribuídos os pesos $\alpha = 1$, $\beta = 0,5$ e $\gamma = 0,25$, respectivamente. O módulo responsável pelo cálculo da medida permite contudo parametrizar não só o valor dos pesos mas também o valor dado à identificação (ver secção 5.6.6). Note-se, aliás, que nada impede que os valores α , β e γ sejam maiores que um²

5.1.3 Cenários selectivos

À semelhança do que aconteceu no Primeiro HAREM, os sistemas puderam participar num cenário selectivo, ou seja, puderam identificar e classificar apenas um subconjunto das categorias, tipos e subtipos que haviam sido definidos como alvo da avaliação. Assim, além de ser avaliado no cenário total, cada sistema foi avaliado no cenário selectivo que indicou.

O cenário selectivo, apesar de ter a vantagem de permitir avaliar um sistema no cenário que se propôs realizar, tem a limitação de não permitir comparar directamente sistemas que participem em cenários diferentes. Por isso, no Segundo HAREM, fomos mais longe e avaliámos cada participação não só no cenário total e no cenário que o participante propôs, mas também em função dos cenários selectivos propostos pelos outros participantes. Além disso, os sistemas foram ainda avaliados em cenários constituídos apenas por cada uma das categorias previstas nas directivas, como foi feito no Primeiro HAREM. Esta opção permitiu, portanto, uma comparação mais fina de cada sistema com os restantes.

Definimos, assim, os seguintes tipos de cenário:

- **Cenário de avaliação:** conjunto de categorias, tipos e subtipos em que se pretende avaliar uma participação.
- **Cenário de participação:** conjunto de categorias, tipos e subtipos que o sistema se propõe classificar.

Na figura 5.2, ilustra-se de forma muito genérica e simplificada os dois tipos de cenário, e no que consiste a avaliação nesses cenários. Como se pode ver, as categorias reconhecidas por uma dada participação constituem um cenário de participação (não se encontra ilustrado, mas o cenário de participação pode naturalmente ser coincidente com o cenário total). Cada um dos cenários de participação é então usado como cenário de avaliação, levando a que haja um processo de adaptação tanto da CD como das várias participações a esse cenário. Veja-se que na avaliação no cenário selectivo C1, a CD passa a ter menos entidades (todas as entidades cujas categorias não façam parte do cenário são removidas),

² A fórmula proposta não permite anular completamente o peso da identificação. Por isso, e por uma questão de generalidade, estamos agora convencidos de que teria sido melhor também associar um peso independente à identificação.

e as entidades da categoria LOCAL deixam de ter o atributo TIPO; como a participação B só tem entidades da categoria LOCAL, a única adaptação consiste em fazer desaparecer o atributo TIPO. A mesma figura ilustra também que se pode ter como cenário de avaliação, um cenário selectivo (C3) diferente dos cenários de participação.

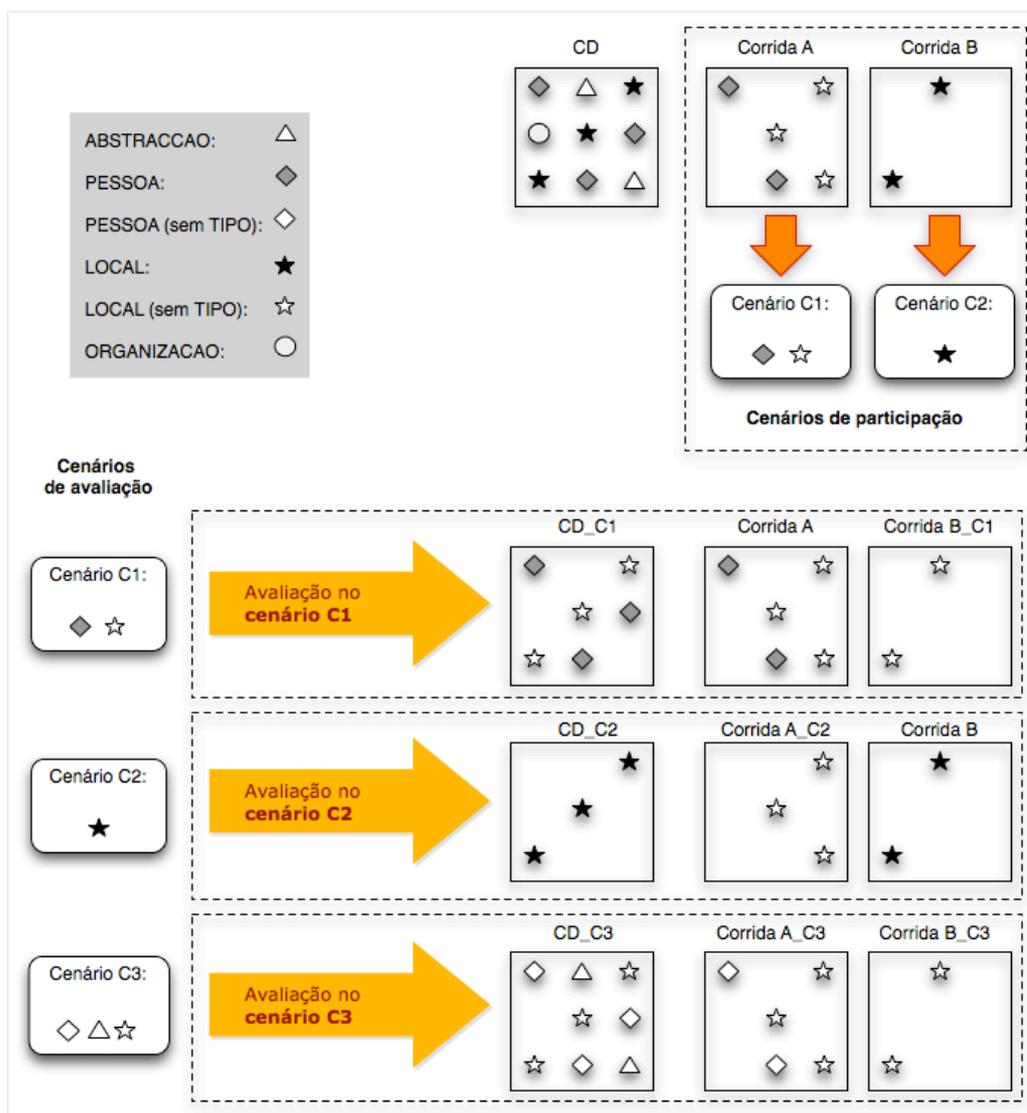


Figura 5.2: Exemplos de cenários de participação e avaliação

Note-se, a este respeito, que, na avaliação num dado cenário selectivo, por oposição ao que aconteceria se o cenário de avaliação fosse o cenário total, a diminuição do número de categorias, tipos ou subtipos irá reflectir-se na medida utilizada (ver secção 5.1.2).

Um problema que se nos pôs relativamente à avaliação por cenários selectivos foi a existência de várias participações com EM que apenas foram identificadas, ou seja, sem qualquer tipo de classificação. Depois de reflexão aturada, decidimos utilizar também o cenário de participação para, nestes casos, tomar a decisão de manter ou não a EM. Quando a EM não tem qualquer classificação assume-se que poderia ter qualquer uma das classificações possíveis dentro do cenário de participação. Sendo assim, a EM é removida apenas se o cenário de participação não tiver classificações comuns com o cenário de avaliação, ou seja, se a intersecção dos cenários for nula. Se isto não se verificar, a EM mantém-se. Alguns exemplos da utilização de ambos os tipos de cenário selectivo encontram-se mais à frente neste capítulo, na figura 5.9.

5.1.4 Avaliação de ALT

A etiqueta ALT foi utilizada para anotar todas as segmentações possíveis de EM (veja-se a secção respectiva do capítulo 1 para uma motivação linguística desta opção, assim como as regras sistemáticas de segmentação de ALT, no apêndice D).

Os participantes foram incentivados a utilizar esta etiqueta e a sua avaliação foi feita de duas formas:

- **Avaliação estrita de ALT**, onde são contabilizadas todas as alternativas possíveis para um segmento de texto, tendo cada alternativa um peso igual ao inverso do número de alternativas dentro desse segmento. Por exemplo, a todos os valores atribuídos a EM dentro de um ALT com três elementos será associado um peso de 1/3. O sistema só terá, assim, o valor máximo possível se tiver apresentado as três alternativas no seu resultado.
- **Avaliação relaxada de ALT**, onde é seleccionado o elemento do ALT que maximiza a classificação do sistema, tal como foi feito no Primeiro HAREM.

5.2 Avaliação da pista do TEMPO

Tal como referimos no capítulo 3, a avaliação da pista do TEMPO foi feita de forma integrada com o HAREM clássico. No entanto, foi necessário ter em conta que as entidades da categoria TEMPO têm atributos específicos, além dos atributos do HAREM clássico.

Na avaliação do TEMPO estendido tivemos por objectivo dar crédito adicional aos sistemas pelo correcto preenchimento dos atributos estendidos da categoria TEMPO, ou seja, TEMPO_REF, SENTIDO, VAL_NORM e VAL_DELTA (cf. capítulo 2), sem os penalizar pela atribuição de valores espúrios. Assim, a fórmula completa usada na avaliação das entidades com a categoria TEMPO encontra-se na figura 5.3. Repare-se que a parcela inicial 5.2 é a medida de avaliação do HAREM clássico apresentada na figura 5.1, sendo as restantes parcelas usadas para avaliar os atributos estendidos de TEMPO.

Destaca-se ainda que, enquanto os atributos VAL_NORM para o tipo DURACAO e VAL_DELTA são avaliados como um todo, não sendo valorizado o correcto preenchimento de cada um dos campos que compõe o valor desses atributos, a situação é diferente para o atributo VAL_NORM quando o tipo é DATA ou HORA. Neste caso, cada um dos campos individuais do atributo contribui separadamente para o valor da medida de classificação da EM. Exemplificando, no

$$1 + \sum_{i=1}^N \left(\left(1 - \frac{1}{n_{cats}}\right) \cdot cat_{certo_i} \cdot \alpha + \left(1 - \frac{1}{n_{tipos}}\right) \cdot tipo_{certo_i} \cdot \beta + \left(1 - \frac{1}{n_{sub}}\right) \cdot sub_{certo_i} \cdot \gamma \right) - \sum_{i=0}^M \left(\frac{1}{n_{cats}} \cdot cat_{esp_i} \cdot \alpha + cat_{certo_i} \cdot \frac{1}{n_{tipos}} \cdot tipo_{esp_i} \cdot \beta + tipo_{certo_i} \cdot \frac{1}{n_{sub}} \cdot sub_{esp_i} \cdot \gamma \right) \quad (5.2)$$

$$+ tr_{certo} \cdot \delta + s_{certo} \cdot \lambda \quad (5.3)$$

$$+ \begin{cases} vd_{certo} \cdot \epsilon \\ vn_{certo} \cdot \epsilon \\ (E_{certo} + A_{certo} + D_{certo} + H_{certo} + M_{certo} + ES_{certo} + lim_{certo}) \cdot \xi \\ (H_{certo} + M_{certo} + lim_{certo}) \cdot \eta \end{cases} \quad (5.4)$$

$$K_{certo_i} = \begin{cases} 1 & \text{se o atributo } K_i \text{ estiver correcto,} \\ 0 & \text{se } K_i \text{ estiver incorrecto ou omisso} \end{cases}$$

$$K_{esp_i} = \begin{cases} 1 - K_{certo_i} & \text{se o atributo } K_i \text{ estiver preenchido} \\ 0 & \text{se } K_i \text{ estiver omisso} \end{cases}$$

$K \in \{cat, tipo, sub\}$

N = número de diferentes classificações vagas na CD, de acordo com o cenário selectivo.

M = número de classificações espúrias na participação, de acordo com o cenário selectivo.

tr, s, vd, vn : classificação referente, respectivamente, aos atributos TEMPO_REF, SENTIDO, VAL_DELTA, e VAL_NORM (quando TIPO="DURACAO").

E, A, D, H, M, ES, lim : classificação referente, respectivamente, aos campos era, ano, dia, mês, hora, minutos, estação e limite do atributo VAL_NORM (quando TIPO="DATA" e TEMPO_REF="ABSOLUTO").

H, M, lim : classificação referente, respectivamente aos campos hora, mês e limite do atributo VAL_NORM (quando TIPO="HORA").

α, β, γ = parâmetros correspondentes aos pesos das categorias, tipos e subtipos.

$\delta, \lambda, \epsilon, \xi, \eta$ = parâmetros correspondentes aos pesos dos atributos estendidos.

Figura 5.3: Fórmula da medida de avaliação do TEMPO estendido

primeiro caso, se o campo H (da hora) estiver mal preenchido, todo o atributo é considerado incorrecto, mas, no segundo caso, o facto de esse campo estar mal preenchido não impede que os restantes sejam considerados certos e pontuados como tal.

Esta distinção pareceu-nos, por um lado, modelar melhor o que se espera com o procedimento de normalização, e, pelo outro, nivelar de forma mais apropriada os sistemas. Por exemplo, para a frase 5.1, qual dos seguintes valores de VAL_NORM propostos por um sistema participante deverá ter uma medida mais elevada: A0M0S20D2H0M0S0 ou A0M0S3D2H0M0S0?

(5.1) O evento durou <EM ID="tp-1" CATEG="TEMPO" TIPO="DURACAO" VAL_NORM="A0M0S23D2H0M0S0">vinte e três semanas e 2 dias

De facto, A0M0S20D2H0M0S0 (correspondente a vinte semanas e dois dias) está mais próximo do valor pretendido do que A0M0S3D2H0M0S0 (três semanas e dois dias), mas não tratámos estes dois casos de forma diferente porque muito possivelmente uma penalização realista teria de depender da aplicação prática do sistema, e podemos facilmente conceber casos em que “mais próximo” é igualmente deficiente.

No entanto, dado que a uma mesma expressão temporal com valor durativo pode corresponder mais do que uma representação de `VAL_NORM` (ou de `VAL_DELTA`, no caso de uma data com valor referencial), o valor desses atributos é considerado como certo se, ao ser convertido para uma unidade mínima, o valor convertido for igual ao valor da CD após conversão para a mesma unidade. Entende-se por unidade mínima a menor unidade usada para especificar este atributo, para a entidade em causa, na CD e na participação. As conversões são feitas de acordo com a tabela que se encontra nas directivas do TEMPO (cf. tabela B.6, no apêndice B).

Continuando com o mesmo exemplo, o valor `A0M5S0D13H0M0S0` (correspondente a cinco meses e treze dias) representa a mesma expressão e seria deste modo considerado como certo, caso fosse fornecido por um sistema em alternativa à resposta `A0M0S23D2H0M0S0`, que estaria na CD. De acordo com a tabela de conversões, que convencionou que um mês são trinta dias, e que uma semana corresponde a sete dias, ambos os valores seriam convertidos para 163 dias antes de serem comparados.

De acordo com a fórmula que apresentámos, os sistemas podem ser avaliados de quatro modos distintos no que diz respeito às entidades da categoria TEMPO:

Clássico, ignorando as parcelas referentes aos atributos estendidos, correspondendo portanto à avaliação dos atributos `CATEG`, `TIPO` e `SUBTIPO`;

Estendido completo, usando todas as parcelas da fórmula;

Estendido sem normalização, usando as parcelas (5.2) e (5.3);

Estendido só com normalização, usando as parcelas (5.2) e (5.4).

De certa forma, estes modos de avaliação na pista do TEMPO são semelhantes à avaliação por cenários selectivos, em que apenas um subconjunto dos atributos é tido em conta.

5.3 Avaliação do ReReLEM

5.3.1 Pontuações e medidas

A avaliação do ReReLEM é feita através da comparação entre as relações que estão na participação e as relações que estão na CD do ReReLEM.

De acordo com essa comparação, cada relação pode receber as pontuações: *Correcta*, *Espúria* ou *Em falta*, sendo que a medida utilizada é igual a 1 se a relação for correcta ou 0 caso contrário.

5.3.2 Expansão de relações

A fim de facilitar a tarefa dos sistemas mas, principalmente, de facilitar o processo de anotação humana, decidimos que não seria necessário anotar **todas** as relações entre EM que existem num documento.

Porém, para poder fazer a avaliação é necessário explicitar tanto na CD como na participação todas as relações que se encontram implícitas. Para tanto, essas relações são expandidas de forma automática pelo módulo *Expandidor* (ver em mais detalhe a sua descrição na secção 5.8.2).

A expansão consiste na explicitação de relações inversas (no caso de existirem), seguida pela aplicação de regras de expansão a pares de relações (como já mencionado no capítulo 4), o que pode dar origem a novas relações, antes apenas implícitas no texto.

5.3.3 Selecção de alinhamentos

A avaliação do ReRelEM centra-se apenas na avaliação de relações, e não na de EM, que é da responsabilidade do HAREM clássico. Assim, a avaliação de cada participação no ReRelEM é feita com base no subconjunto de anotações que são comuns à CD, obtido através do seguinte procedimento:

- remoção, na CD, de todas as EM em falta e respectivas relações;
- remoção, na participação, de todas as EM espúrias e respectivas relações;
- remoção de todas as classificações de EM diferentes do lado da CD e da participação, mantendo apenas as classificações comuns;
- remoção das relações entre facetas que tiverem sido removidas no ponto anterior (ver capítulo 4 para a noção de faceta).

5.4 Métricas

Nas secções anteriores foram apresentados os princípios, as pontuações e as medidas utilizadas na avaliação das três pistas integradas no Segundo HAREM. Quanto às métricas utilizadas, estas foram sempre:

- **Precisão:** afere a qualidade da participação, em termos da proporção de respostas correctas dentro do total de respostas dadas.
- **Abrangência:** afere a qualidade da participação, em termos da proporção de respostas correctas no universo de respostas possíveis.
- **Medida F:** combina a precisão e a abrangência para cada tarefa, de acordo com a seguinte fórmula:

$$Medida F = \frac{2 \cdot Precisão \cdot Abrangência}{Precisão + Abrangência} \quad (5.5)$$

A noção de resposta correcta varia naturalmente em função da medida utilizada. Concretizando e resumindo, para o HAREM clássico a unidade é o valor da medida já apresentada, para o TEMPO estendido é o valor das medidas anteriormente apresentadas e para o ReRelEM é o número de relações correctas.

5.5 Vista geral da arquitectura

Esta secção apresenta a arquitectura da plataforma utilizada na avaliação do Segundo HAREM: uma arquitectura modular, fortemente apoiada na arquitectura de avaliação do Primeiro HAREM (Seco et al. (2007)), onde se procurou que existissem vários módulos, cada um com a função de executar uma tarefa simples e específica.

5.5.1 Formato das colecções

No Segundo HAREM, seguindo as sugestões de [Martins e Silva \(2007\)](#) e de [Almeida \(2007\)](#), optou-se por utilizar colecções de documentos na notação XML, tendo como principal objectivo facilitar o processamento e a validação do material.

Uma colecção é então constituída por vários documentos delimitados pela etiqueta `DOC`. Essa etiqueta contém um atributo, `DOCID`, que é preenchido com um identificador único do documento na colecção. O documento pode conter outras etiquetas XML, como acontece com os textos anotados:

- Todas as entidades mencionadas são anotadas com a etiqueta `EM`, tendo ainda atributos para a sua classificação (`CATEG`, `TIPO` e `SUBTIPO`), atributos para identificar as suas relações com outras `EM` (`COREL` e `TIPOREL`) e atributos específicos do `TEMPO` (`TEMPO_REF`, `SENTIDO`, `VAL_NORM`, e `VAL_DELTA`).
- O elemento `ALT` representa um conjunto de análises alternativas de identificação (onde podem estar incluídas `EM`), separadas pelo carácter “|”.

A correcção do formato dos vários ficheiros é verificado por meio de um validador, criado para o efeito.³

5.5.2 Os módulos

Todos os módulos foram implementados em Java, com a excepção do Gerador de relatórios individuais (implementado em R) e do Avaliador do `TEMPO` estendido (implementado em Awk). A biblioteca Java JDOM⁴ foi utilizada para a manipulação de XML.

Cada módulo tem como entrada um (ou mais) ficheiros de texto, que são processados de forma a produzir um resultado pronto a ser tratado pelo módulo seguinte. A figura 5.4 representa todos os módulos de avaliação do Segundo HAREM, na ordem pela qual devem ser utilizados para se chegar aos resultados da avaliação. Podem ver-se não só os módulos para a avaliação do HAREM clássico mas também a indicação de onde entram os módulos para a avaliação do `TEMPO` estendido e do `ReReEM`. A parte do diagrama com o fundo verde corresponde à avaliação individual de uma participação num determinado cenário selectivo; os módulos que se encontram fora desse fundo verde (ou seja, o Gerador de resultados HAREM e o Gerador de relatórios individuais), utilizam como entrada o conjunto de todas as avaliações.

Há ainda que referir o ficheiro `harem.conf`, que contém toda a informação relativa às possibilidades válidas para o preenchimento dos atributos das `EM`:

- Árvore de categorias, tipos e subtipos;
- Atributos específicos do `TEMPO`;
- Tipos de relação e respectivas inversas, se existirem.

Nas próximas secções encontra-se uma descrição mais detalhada dos vários módulos.

³ Desenvolvido inicialmente por David Cruz e melhorado por Luís Miguel Cabral e mais tarde integrado no SAHARA (ver apêndice G).

⁴ <http://www.jdom.org>

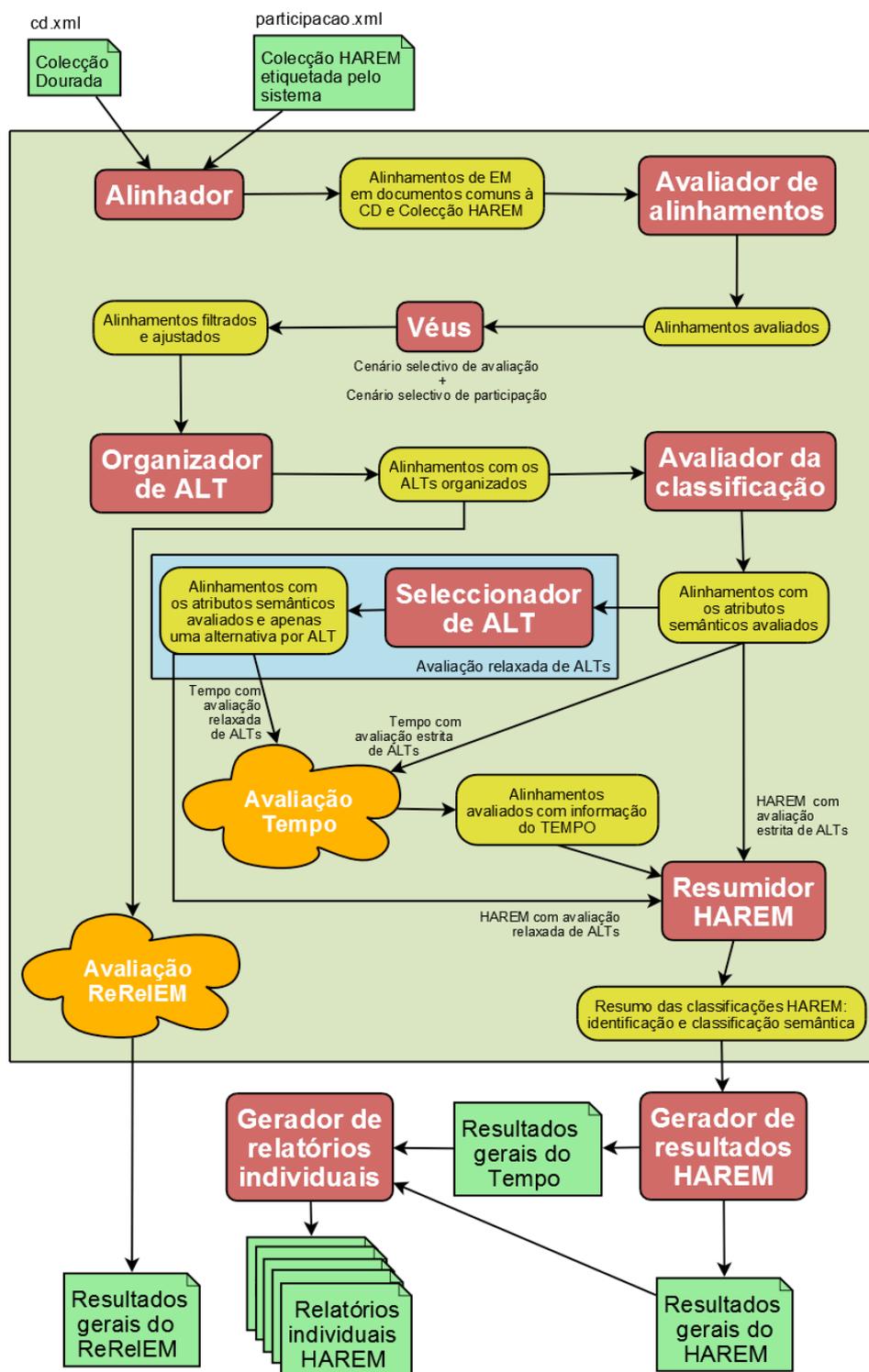


Figura 5.4: Arquitectura dos programas de avaliação no Segundo HAREM

```
#comentários
DOC DOCID\_1
<VERIFICACAO\_MANUAL>Informação para o juiz humano</VERIFICACAO\_MANUAL>
Alinhamento 1
Alinhamento 2
(...)
Alinhamento n
DOC DOCID\_2
(...)
DOC DOCID\_N
(...)
```

Figura 5.5: Formato da saída do Alinhador

5.6 Módulos de avaliação do HAREM clássico

Como já foi dito, grande parte dos módulos utilizados na avaliação do HAREM clássico foi construída a partir de módulos já criados para o Primeiro HAREM (Seco et al. (2007)), mais propriamente o Alinhador, o Avaliador de alinhamentos, o Véus, o Avaliador da classificação, o Resumidor das classificações e o Gerador de resultados⁵. Em alguns casos, acabaram contudo por existir alterações substanciais devido a todas as novidades do Segundo HAREM.

5.6.1 Alinhador

Uma participação é alinhada com a CD utilizando o Alinhador.

Este módulo alinha EM de documentos que estão tanto na CD como na participação, ignorando os outros documentos da Coleção do Segundo HAREM que estejam presentes na participação. Desta forma não é necessário utilizar um programa anterior para extrair da participação os documentos que fazem parte da CD, como acontecia no Primeiro HAREM.

5.6.1.1 Formato da saída

A saída do Alinhador é exemplificada na figura 5.5. As primeiras linhas, iniciadas por #, podem conter comentários, como por exemplo o nome da CD e participação a que se refere o alinhamento. De seguida, o início dos alinhamentos de cada documento é identificado através de uma linha iniciada por DOC, seguida pelo identificador do documento (valor do elemento DOCID).

Os alinhamentos podem ser de um dos cinco tipos que existiam no Primeiro HAREM conforme o exemplo adaptado de Seco et al. (2007):

- **um para um:** uma EM da CD alinha exactamente a uma EM na participação.
`17:00 --> [17:00]`
- **um para muitos:** uma EM da CD alinha a mais do que uma EM na participação.
`17:00 --> [17, 00]`

⁵ Estes módulos eram designados no Primeiro HAREM AlinhEM, AvalIDa, Véus, Emir, Ida e Sultão, respectivamente.

```

<EM>17:00<EM> ----> [<EM>17:00</EM>]: [Correcto]}
<EM>17:00<EM> ----> [<EM>17</EM>, <EM>00</EM>]: [Parcialmente\_Correcto\_por\_Defeito(0.25;
0.75), Parcialmente\_Correcto\_por\_Defeito(0.25; 0.75)]}

```

Figura 5.6: Avaliação de alinhamentos

- **muitos para um:** mais do que uma EM da CD alinham a uma EM na participação.
`17 --> [17:00]`
`00 --> [17:00]`
- **nenhum para um:** uma EM é identificada na participação mas não há uma EM correspondente na CD. `<EM CATEG="ESPURIO">Ontem --> [Ontem]`
- **um para nenhum:** uma EM da CD não foi marcada como tal na participação.
`Departamento de Informática --> [null]`

5.6.1.2 Etiquetas ALT

O tratamento das etiquetas ALT sofreu uma alteração considerável no Segundo HAREM: agora, estas etiquetas podem surgir tanto na CD como nas saídas dos participantes, que foram incentivados a utilizá-las para marcar conjuntos de análises alternativas de identificação. A sua utilização foi, portanto, também alvo de avaliação, o que levou a que o Alinhador tivesse de estar preparado para lidar com participações que enviassem mais do que uma segmentação para os mesmos fragmentos de texto.

5.6.1.3 Etiquetas OMITIDO

As etiquetas OMITIDO foram utilizadas pelas mesmas razões que o foram no Primeiro HAREM, nomeadamente para permitir algum controlo sobre partes do texto sem relevância linguística para a avaliação conjunta Santos e Cardoso (2007b). Devido à adopção do XML, os programas reconhecem agora OMITIDO como um elemento válido (embora apenas fazendo sentido do lado da CD). Internamente, sempre que um bloco de texto ocorre dentro de etiquetas OMITIDO, esse bloco é tratado como se fosse uma EM, com um atributo particular, que indica que é para ser omitida (OMITIDO). Ao escrever os alinhamentos na saída, o Alinhador ignora todos aqueles que do lado da CD tiverem uma EM com o atributo OMITIDO.

5.6.2 Avaliador de alinhamentos

O Avaliador de alinhamentos é exactamente igual ao utilizado no Primeiro HAREM. Os alinhamentos produzidos pelo Alinhador são avaliados, comparando a delimitação das EM do lado da CD com a delimitação das EM do lado da participação. A classificação é colocada à frente de cada alinhamento, como ilustrado na figura 5.6:

As possíveis pontuações são: Correcto, Em Falta, Espúrio, Parcialmente_Correcto_por_Defeito e Parcialmente_Correcto_por_Excesso. Apesar de no Segundo HAREM não serem contabilizadas EM parcialmente identificadas, o Avaliador mantém estes alinhamentos, que são tratados na fase seguinte.

```

Antes:
<EM>17:00</EM> ----> [<EM>17</EM>, <EM>00</EM>]: [Parcialmente\_Correcto\_por\_Defeito(0.25;
0.75), Parcialmente\_Correcto\_por\_Defeito(0.25; 0.75)]

Depois:
<EM>17:00</EM> ----> [null]: [Em Falta]
<EM CATEG="ESPURIO">17</EM> ----> [<EM>17</EM>]: [Espurio]
<EM CATEG="ESPURIO">00</EM> ----> [<EM>00</EM>]: [Espurio]

```

Figura 5.7: Passagem de entidades parcialmente correcta a espúrias

5.6.3 Véus

O módulo Véus aplica os filtros utilizados para adaptar um conjunto de alinhamentos a um cenário selectivo definido por um conjunto de categorias, tipos e subtipos. Como já explicado na secção 5.1.3, é assim possível avaliar uma participação no cenário que for pretendido (o chamado cenário de avaliação), independentemente de ter sido o cenário em que o sistema participou, um subconjunto desse cenário ou mesmo um cenário que incluía algumas categorias em que o sistema não participou.

Ao contrário do Véus do Primeiro HAREM, o Véus no Segundo HAREM não filtra os documentos por género textual ou variante (PT, BR, etc.), visto que essa informação deixou de estar incluída no cabeçalho dos próprios documentos, passando antes a estar compreendida num ficheiro separado, que designámos como *Meta*.

O Véus do Primeiro HAREM permitia além disso parameterizar a avaliação de acordo com três estilos diferentes, sendo um deles o estilo *mic*, em que os casos parcialmente correctos não eram contabilizados. Esse estilo foi o único utilizado no Segundo HAREM, levando a que, a partir desta fase, deixassem de existir alinhamentos do tipo “*um para muitos*”. As EM parcialmente correctas passaram assim a ser consideradas espúrias e as EM alinhadas com mais de uma EM passaram a contar como estando em falta. Na figura 5.7 ilustra-se a transformação de uma entidade parcialmente correcta em duas entidades espúrias.

5.6.3.1 Representação dos cenários selectivos

O Véus ajusta as EM nos alinhamentos de acordo com ambos os cenários selectivos – o de participação (opção *-sistema*) e o de avaliação (opção *-avaliacao*) –, evitando assim que tenha de ser criada uma CD e uma participação para cada cenário que se pretenda avaliar. Tal já foi ilustrado na anterior figura 5.2.

Os cenários são representados por uma lista com as categorias, tipos e subtipos que estão incluídos nesse cenário. A figura 5.8 ilustra a representação de alguns cenários.

5.6.3.2 Formato da saída

A saída do Véus continua a ser um conjunto de alinhamentos, agora conforme o estilo *mic* e de acordo com os cenários de avaliação e de participação. A única diferença no formato é que na primeira linha da saída é colocada a representação do cenário de avaliação, iniciada pelo carácter #.

Filtro	Descrição
"*(ou sem a opção do cenário)"	Cenário total.
"PESSOA(*) : LOCAL(*) : ORGANIZACAO(*)"	Apenas as categorias PESSOA, LOCAL e ORGANIZACAO e o seu conjunto normal de tipos e subtipos.
"LOCAL(FISICO{*}; HUMANO{*})"	A categoria LOCAL apenas com os tipos FISICO e HUMANO e o seu conjunto normal de subtipos.
"LOCAL(FISICO{*}; HUMANO{RUA, PAIS, DIVISAO, REGIAO})"	A categoria LOCAL apenas com o tipo FISICO como seu conjunto normal de subtipos e com o tipo HUMANO apenas com os subtipos RUA, PAIS, DIVISAO e REGIAO.

Figura 5.8: Exemplos de representação de cenários.

5.6.3.3 Exemplo de aplicação de filtros pelo Véus

A figura 5.9 exemplifica a aplicação de um filtro para avaliar apenas a categoria PESSOA numa participação que concorreu exclusivamente nas categorias PESSOA e ORGANIZACAO. O cenário de participação é utilizado apenas na decisão de manter a situação da última linha da figura, em que existe uma EM não classificada que é espúria, mas que não é eliminada porque tanto o cenário de avaliação como o cenário de participação contêm a categoria PESSOA.

5.6.4 Organizador de ALT

Como já foi dito na secção 5.1.4, no Segundo HAREM existem dois tipos de avaliação relacionados com a utilização de ALT, sendo o Organizador de ALT o módulo que possibilita a avaliação estrita. Esta avaliação obedece aos seguintes três passos:

1. O Alinhador associa as EM da participação às EM na CD, estando estas dentro ou fora de ALT. A saída fica organizada de acordo com os ALT na CD.
2. O Véus remove todos os alinhamentos parcialmente correctos. A partir deste passo, só não existem EM espúrias ou em falta dentro de um ALT se a participação contiver um ALT exactamente igual ao que está na CD. (A aplicação de um filtro para avaliar um cenário selectivo pode também dar origem a alternativas vazias ou duplicadas.)
3. Finalmente, o Organizador de ALT garante que:
 - não haja análises iguais, que possam ter surgido após a aplicação do Véus, por meio da remoção de duplicados;
 - não haja alternativas que deixaram de conter EM após a aplicação do Véus;

Antes	Depois
<EM CATEG="PESSOA"> -> <EM CATEG="PESSOA">	<EM CATEG="PESSOA"> -> <EM CATEG="PESSOA">
<EM CATEG="PESSOA ORGANIZACAO"> -> <EM CATEG="PESSOA">	<EM CATEG="PESSOA"> -> <EM CATEG="PESSOA">
<EM CATEG="PESSOA"> -> <EM CATEG="PESSOA ORGANIZACAO">	<EM CATEG="PESSOA"> -> <EM CATEG="PESSOA">
<EM CATEG="PESSOA"> -> <EM CATEG="ORGANIZACAO">	<EM CATEG="PESSOA"> -> [null]
<EM CATEG="ORGANIZACAO"> -> <EM CATEG="ORGANIZACAO">	nada
<EM CATEG="ORGANIZACAO"> -> [null]	nada
<EM CATEG="ESPURIO"> -> <EM CATEG="ORGANIZACAO">	nada
<EM CATEG="PESSOA"> -> [null]	<EM CATEG="PESSOA"> -> [null]
<EM CATEG="ESPURIO"> -> 	<EM CATEG="ESPURIO"> ->

Figura 5.9: Exemplos da aplicação do Véus, com cenário de participação constituído por PESSOA e ORGANIZACAO e cenário de avaliação PESSOA.

- o peso $1/(\text{total de elementos})$ seja atribuído a cada EM dentro de um ALT (esse peso é colocado no fim de cada alinhamento, depois do caracter '^').

A figura 5.10 exemplifica a evolução de processamento de um ALT ao longo dos passos descritos acima.

5.6.5 Listador de espúrios

O Listador de espúrios é um módulo que pode ser aplicado sobre a saída do Organizador de ALT no caso de se pretender listar todas as EM espúrias, para efeitos de depuração.

5.6.6 Avaliador da classificação

O Avaliador da classificação recebe a saída do Organizador de ALT e calcula a pontuação de cada atributo, comparando as categorias, tipos e subtipos do lado da CD com os mesmos atributos do lado da participação.

A medida da classificação é também calculada aplicando aos atributos de cada EM e respectivas pontuações a fórmula que se encontra na figura 5.1.

5.6.6.1 Formato da saída

A saída do Avaliador da classificação continua a conter os alinhamentos, cuja avaliação (anexada ao final da linha correspondente a cada um) é constituída por:

- Pontuação dos atributos: pontuação do preenchimento de cada atributo nas EM da participação. O resultado da comparação com os atributos das EM na CD pode ser

```

CD:
O <ALT>
<EM ID="Xyz-60" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França de 2009</EM> |
<EM ID="Xyz-60-aa" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França</EM> <EM ID="Xyz-5"
  CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> | <EM ID="Xyz-60-aaa" CATEG
  ="ACONTECIMENTO" TIPO="ORGANIZADO">Tour</EM> de <EM ID="Xyz-61" CATEG="LOCAL" TIPO="
  HUMANO" SUBTIPO="PAIS">França</EM> <EM ID="Xyz-5-aa" CATEG="TEMPO" TIPO="TEMPO_CALEND"
  SUBTIPO="DATA">de 2009</EM>
</ALT> ...

Participação:
O Tour de <EM ID="Xyz_1" CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM> <EM ID="
  Xyz_2" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> ...

Saída do alinhador:
<ALT>
<ALT1>
<EM ID="Xyz-60" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França de 2009</EM> ----> [<EM
  ID="Xyz_1" CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM>, <EM ID="Xyz_2" CATEG
  ="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM>]
</ALT1>
<ALT2>
<EM ID="Xyz-60-aa" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França</EM> ----> [<EM ID="
  Xyz_1" CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM>]
<EM ID="Xyz-5" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> ----> [<EM ID="
  Xyz_2" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM>]
</ALT2>
<ALT3>
<EM ID="Xyz-60-aaa" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour</EM> ----> [null]
<EM ID="Xyz-61" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">França</EM> ----> [<EM ID="Xyz_1"
  CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM>]
<EM ID="Xyz-5-aa" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> ----> [<EM ID="
  Xyz_2" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM>]
</ALT3>
</ALT>

Saída do Véus (com cenário de avaliação "apenas_LOCAL"):
<ALT>
<ALT1>
<EM CATEG="ESPURIO">França</EM> ----> [<EM ID="Xyz_1" CATEG="LOCAL" TIPO="FISICO" SUBTIPO="
  REGIAO">França</EM>]: [ Espurio ]
</ALT1>
<ALT2>
<EM CATEG="ESPURIO">França</EM> ----> [<EM ID="Xyz_1" CATEG="LOCAL" TIPO="FISICO" SUBTIPO="
  REGIAO">França</EM>]: [ Espurio ]
</ALT2>
<ALT3>
<EM ID="Xyz-61" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">França</EM> ----> [<EM ID="Xyz_1"
  CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM>]: [ Correcto ]
</ALT3>
</ALT>

Saída do Organizador de ALT:
<ALT>
<ALT1>
<EM ID="Xyz-61" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">França</EM> ----> [<EM ID="Xyz_1"
  CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM>]: [ Correcto ]^1
</ALT1>
</ALT>

```

Figura 5.10: Evolução de um ALT

```

<EM ID="551" CATEG="ACONTECIMENTO" TIPO="EVENTO">Proclamação da Carta dos Direitos
Fundamentais</EM> ----> [<EM ID="133" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Proclamação
da Carta dos Direitos Fundamentais</EM>]: [{ Categoria(Correcto:[ACONTECIMENTO] Espurio:[ ]
Em_Falta:[ ] Tipo(Correcto:[ ] Espurio:[ORGANIZADO] Em_Falta:[EVENTO]) Subtipo(Correcto:
[ ] Espurio:[ ] Em_Falta:[ ] MaxCSC_CD(2.275) MaxCSC_S(2.275) CSC(1.775) Peso(1.0) }}

<ALT>
<ALT1>
<EM ID="113" CATEG="PESSOA" TIPO="INDIVIDUAL">Filipe II de Espanha</EM> ----> [ null ]: [{
Categoria(Correcto:[ ] Espurio:[ ] Em_Falta:[ ] Tipo(Correcto:[ ] Espurio:[ ] Em_Falta:[ ]
Subtipo(Correcto:[ ] Espurio:[ ] Em_Falta:[ ] MaxCSC_CD(2.3375) MaxCSC_S(0.0) CSC(0.0) Peso
(0.0) PALT(0.5) }}

</ALT1>
<ALT2>
<EM ID="113aa" CATEG="PESSOA" TIPO="INDIVIDUAL">Filipe II</EM> ----> [ null ]: [{ Categoria(
Correcto:[ ] Espurio:[ ] Em_Falta:[ ] Tipo(Correcto:[ ] Espurio:[ ] Em_Falta:[ ] Subtipo(
Correcto:[ ] Espurio:[ ] Em_Falta:[ ] MaxCSC_CD(2.3375) MaxCSC_S(0.0) CSC(0.0) Peso(0.0)
PALT(0.5) }}

<EM ID="2" CATEG="LOCAL|ORGANIZACAO" TIPO="HUMANO|ADMINISTRACAO" SUBTIPO="PAIS|">Espanha</EM>
----> [<EM ID="1629" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="DIVISAO">Espanha</EM>]: [{
Categoria(Correcto:[LOCAL] Espurio:[ ] Em_Falta:[ORGANIZACAO]) Tipo(Correcto:[HUMANO]
Espurio:[ ] Em_Falta:[ ] Subtipo(Correcto:[ ] Espurio:[DIVISAO] Em_Falta:[PAIS]) MaxCSC_CD
(3.758333333333333) MaxCSC_S(2.4833333333333334) CSC(2.2333333333333334) Peso(1.0) PALT
(0.5) }}

</ALT2>
</ALT>

<EM ID="185" _CATEG="PESSOA" _TIPO="INDIVIDUAL">Sócrates </EM> ----> [ null ]: [{ Categoria(Correcto:
[ ] _Espurio:[ ] _Em_Falta:[ ])_Tipo(Correcto:[ ] _Espurio:[ ] _Em_Falta:[ ])_Subtipo(Correcto:[ ] _
Espurio:[ ] _Em_Falta:[ ])_MaxCSC_CD(2.3375)_MaxCSC_S(0.0)_CSC(0.0)_Peso(0.0) }}

<EM _CATEG="ESPURIO">Escola </EM> ----> [<EM ID="111" _CATEG="LOCAL" _TIPO="HUMANO" _SUBTIPO="
DIVISAO">Escola </EM>]: [{ Categoria(Correcto:[ ] _Espurio:[ ] _Em_Falta:[ ])_Tipo(Correcto:[ ] _
Espurio:[ ] _Em_Falta:[ ])_Subtipo(Correcto:[ ] _Espurio:[ ] _Em_Falta:[ ])_MaxCSC_CD(0.0)_
MaxCSC_S(2.4833333333333334)_CSC(0.0)_Peso(0.0) }}

```

Figura 5.11: Exemplos de alinhamentos após a aplicação do Avaliador da classificação.

pontuado como: Correcto, Espúrio ou Em Falta. Se uma categoria não estiver correcta, o tipo e o subtipo não são pontuados e, se um tipo não estiver correcto, o subtipo também não é pontuado.

- CSC: classificação efectiva da anotação do sistema, comparada com a anotação da CD;
- Peso: peso do alinhamento (0 se for Espúrio ou Em Falta, 1 se estiver Correcto). No Primeiro HAREM poderia ter valores entre 0 e 1 para entidades parcialmente identificadas.
- PALT: peso do alinhamento dentro de um ALT (por omissão é igual a 1);
- MaxCSC_CD: classificação máxima da anotação na CD;
- MaxCSC_S: classificação máxima possível com a anotação presente na participação.

Alguns exemplos de alinhamentos após a aplicação do Avaliador da classificação no cenário total estão na figura 5.11.

```
Avaliação Global – Classificação
Valor máximo possível para a Classificação na CD: 15971.775396825311
Valor máximo possível para a Classificação do sistema: 11681.552480158529
Valor da Classificação do sistema: 7424.880753968094
Precisão Máxima do Sistema: 0.6356073618279316
Abrangência Máxima na CD: 0.4648751043321037
Medida F: 0.5369972675258301
```

Figura 5.12: Fragmento de saída do resumidor de classificações

5.6.7 Seleccionador de ALT

O Seleccionador de ALT é utilizado apenas na avaliação relaxada de ALT. É aplicado sobre a saída do Avaliador da classificação e tem como objectivo seleccionar apenas o melhor elemento dentro de cada ALT.

Existem dois critérios para a selecção do melhor elemento, empregues em sequência (ou seja, se o primeiro não permitir decidir, recorre-se ao segundo):

1. Elemento com a melhor medida F.
2. Elemento com com o maior valor da medida de classificação.

5.6.8 Resumidor das classificações

O Resumidor das classificações processa a saída do Avaliador da classificação, utilizando os valores que este último calcula para cada alinhamento para obter as medidas totais na participação e aferir o desempenho do sistema. São apresentadas as medidas tanto para a tarefa de classificação como para identificação apenas (onde é considerado um ponto por EM correctamente identificada – por outras palavras, na fórmula de avaliação apresentada na figura 5.1, os pesos α , β e γ são zero).

A figura 5.12 mostra um fragmento da saída do Resumidor das classificações, onde se pode ver o valor da precisão, da abrangência e da medida F obtidos por um sistema. Há ainda a referir que no cabeçalho da saída do Resumidor das classificações se encontra a representação do cenário selectivo em questão.

5.6.9 Gerador de resultados

O Gerador de resultados tem como objectivo juntar os resultados de todas as saídas do Resumidor de classificações, comparando-os de forma a criar uma tabela (em HTML) com as participações ordenadas pela melhor medida F, mostrando ainda outros valores, como a precisão e a abrangência.

O Gerador de resultados é invocado uma vez para cada cenário. Dessa forma, é gerada uma tabela para cada um dos cenários avaliados, cuja consulta facilita a visualização e comparação do desempenho dos vários sistemas.

5.6.10 Gerador de relatórios individuais

O Gerador de relatórios individuais gera um relatório de desempenho detalhado por cada sistema participante. Este módulo processa os ficheiros em HTML produzidos

pelo Gerador de resultados para a avaliação do HAREM clássico e para a avaliação do TEMPO estendido, mas não para a avaliação do ReReLEM.

Cada relatório inclui tabelas e gráficos individuais de um sistema (ou seja, contendo apenas os resultados das suas corridas). Por exemplo, na figura 5.13(a) ilustra-se o gráfico comparativo dos cenários de avaliação com o desempenho das corridas do sistema Cage2.

O relatório inclui igualmente gráficos comparativos com todas as corridas avaliadas no Segundo HAREM, de forma a que as corridas possam ser comparadas como se fossem de sistemas diferentes (embora as corridas do sistema a que o relatório diz respeito se encontrem destacadas). Veja-se, por exemplo, a figura 5.13(b), que mostra o gráfico precisão/abrangência gerado para o relatório do sistema da Priberam, e a figura 5.13(c), que mostra o gráfico das três métricas de avaliação gerado para o relatório do sistema REMBRANDT.

Além disso, foram também criados gráficos que mostram o desempenho em termos de medida F das várias corridas agrupadas pelo sistema a que pertencem. Este último gráfico permite comparar de forma mais imediata os sistemas (cf. figura 5.13(d), que destaca o desempenho do sistema XIP-L2F/Xerox).

Os gráficos e tabelas são criados para todos os cenários de avaliação, modos de avaliação do TEMPO estendido e avaliação por categoria.

5.7 Módulo de avaliação da pista do TEMPO

Se as entidades da categoria TEMPO apenas tivessem os atributos do HAREM clássico CATEG, TIPO e SUBTIPO, os sistemas teriam sido avaliados tendo por base apenas a CD do Segundo HAREM e o encadeamento de avaliação do HAREM clássico teria sido suficiente para fazer a avaliação dos sistemas na pista do TEMPO, de acordo com o que foi descrito na secção anterior. Como tal não foi o caso, foi necessário incluir na sequência de processamento um módulo adicional para avaliar as entidades da categoria TEMPO quanto aos atributos estendidos na CD do TEMPO, o que é ilustrado na figura 5.4 com a nuvem *Avaliação TEMPO*.

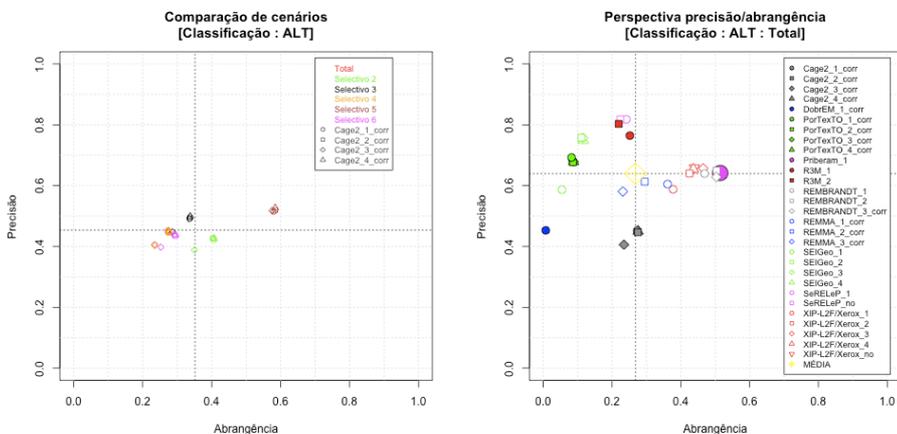
A avaliação dos sistemas na pista do TEMPO usando a CD do TEMPO pressupõe então que a avaliação da classificação, ou seja, o HAREM clássico, tenha sido previamente realizada. Ou seja, o procedimento de avaliação segue o encadeamento esquematizado na figura 5.4, em que *cd.xml*, a entrada do processo, corresponde à CD do TEMPO.

A nuvem de avaliação do TEMPO estendido contém apenas um módulo que analisa os alinhamentos após terem sido processados pelo módulo Avaliador da classificação e produz um novo ficheiro de alinhamentos. Cada linha nesse ficheiro corresponde a uma linha do ficheiro original, mas modificada para conter a avaliação dos atributos estendidos de TEMPO. Essa avaliação é feita de acordo com a fórmula descrita na secção 5.2.

A figura 5.14 mostra uma linha no ficheiro de entrada, que contém o valor da medida de classificação para a entidade em análise, e a linha correspondente no ficheiro de saída, com a informação específica da avaliação dos atributos adicionais de TEMPO.

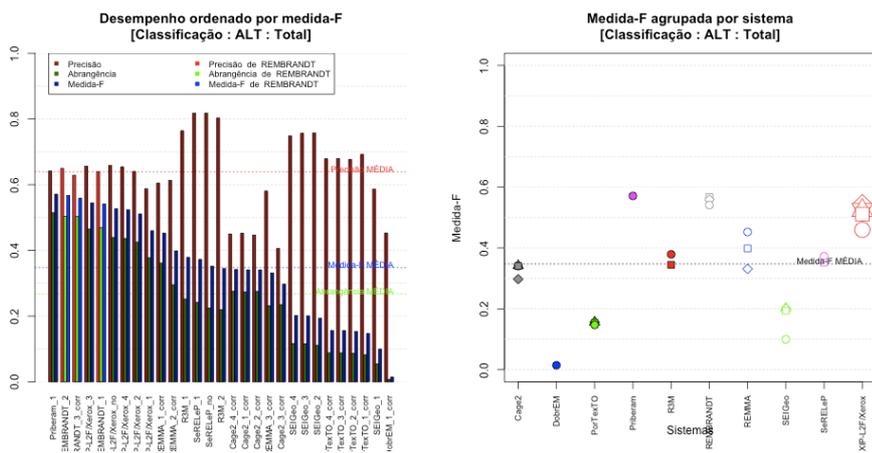
O módulo de avaliação do TEMPO estendido analisa cada alinhamento do ficheiro de entrada, produzindo um novo, em que os valores $MaxCSC_CD$, $MaxCSC_S$ e CSC são incrementados com o valor adicional máximo do lado da CD, do lado do sistema e o valor acrescentado que o sistema tem ao acertar nos atributos estendidos, respectivamente.

O alinhamento passa também a incluir três campos auxiliares que indicam a decomposição dos valores totais. No exemplo, estes campos estão preenchidos com os seguintes



(a) Gráfico de precisão/abrangência comparativo dos vários cenários para um único participante

(b) Gráfico de precisão/abrangência comparativo das várias corridas participantes, com um dos sistemas destacados



(c) Gráfico de precisão, abrangência e medida F comparativo das várias corridas participantes, com um dos sistemas destacados

(d) Gráfico de medida F comparativo das várias corridas das participantes agrupadas por sistema, com um dos sistemas destacados

Figura 5.13: Exemplos de gráficos presentes nos relatórios individuais

valores:

$$\begin{aligned} \text{MaxCSC_CD} &= \text{MaxCSC_CD_class}(2.4875) + \text{MaxCSC_CD_TEMPO}(1) \\ \text{MaxCSC_S} &= \text{MaxCSC_S_class}(2.4875) + \text{MaxCSC_S_TEMPO}(1) \\ \text{CSC} &= \text{CSC_class}(2.4875) + \text{CSC_TEMPO}(1). \end{aligned}$$

Além disso, são igualmente adicionados campos que indicam se os atributos estão correctos ou em falta.

```

Entrada:
<EM ID="H2-Ren_2003_6465-168" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="HORA" VAL_NORM="
+-----T1806E-LM-">às 18h06</EM> ----> [<EM ID="251-26003-26011" CATEG="TEMPO" TIPO="
TEMPO_CALEND" SUBTIPO="HORA" VAL_NORM="+-----T1806E-LM-">às 18h06</EM>]::{{ Categoria(
Correcto:[TEMPO] Espurio:[ ] Em_Falta:[ ] Tipo(Correcto:[TEMPO_CALEND] Espurio:[ ]
Em_Falta:[ ] Subtipo(Correcto:[HORA] Espurio:[ ] Em_Falta:[ ] MaxCSC_CD(2.4875) MaxCSC_S
(2.4875) CSC(2.4875) Peso(1.0) }}

Saída:
<EM ID="H2-Ren_2003_6465-168" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="HORA" VAL_NORM="
+-----T1806E-LM-">às 18h06</EM> ----> [<EM ID="251-26003-26011" CATEG="TEMPO" TIPO="
TEMPO_CALEND" SUBTIPO="HORA" VAL_NORM="+-----T1806E-LM-">às 18h06</EM>]::{{ Categoria(
Correcto:[TEMPO] Espurio:[ ] Em_Falta:[ ] Tipo(Correcto:[TEMPO_CALEND] Espurio:[ ]
Em_Falta:[ ] Subtipo(Correcto:[HORA] Espurio:[ ] Em_Falta:[ ] MaxCSC_CD(3.4875) MaxCSC_S
(3.4875) CSC(3.4875) Peso(1.0) }}:{{ ValNorm(Correcto:[+-----T1806E-LM-] Em_Falta:[ ]
MaxCSC_CD=MaxCSC_CD_class(2.4875)+MaxCSC_CD_TEMPO(1) MaxCSC_S=MaxCSC_S_class(2.4875)+
MaxCSC_S_TEMPO(1) CSC=CSC_class(2.4875)+CSC_TEMPO(1) }}

```

Figura 5.14: Avaliação dos atributos estendidos de entidades da categoria TEMPO

Após esta actualização dos alinhamentos, o ficheiro será processado pelo módulo Resumidor de classificações, seguindo o processamento normal que seguiria a avaliação do HAREM clássico, de modo a produzir as medidas de precisão, abrangência e medida F no preenchimento dos atributos do TEMPO.

5.8 Módulos de avaliação do ReReEM

Esta secção apresenta a avaliação da pista do ReReEM, que consiste em reconhecer relações entre EM (ver capítulo 4), e descreve também uma aplicação criada para visualizar grafos onde estejam representadas as relações anotadas num documento com o esquema de anotação do ReReEM (ver secção 5.8.9).

Como se pode verificar na figura 5.15, a primeira etapa da avaliação é a conversão da notação da CD do ReReEM e das participações para uma notação compatível com a esperada pelos módulos (ver secção 5.8.1). A fase seguinte utiliza alguns dos módulos desenvolvidos para o HAREM clássico e, por fim, acontece a avaliação de relações, que compreende a expansão das relações, a selecção dos alinhamentos sujeitos à avaliação, a normalização dos identificadores das EM da participação, a alteração da representação utilizada nos ficheiros intermédios para focar as relações, a filtragem de alguns tipos de relação e a avaliação propriamente dita, com a geração de tabelas com os resultados.

5.8.1 Conversão de notação

A conversão de notação é a primeira etapa da avaliação do ReReEM, realizada antes mesmo da CD do ReReEM e da participação serem processadas pelos módulos do HAREM clássico. Esta conversão é necessária uma vez que foi criada uma nova notação para possibilitar a identificação e facilitar a leitura de relações entre diferentes facetas de EM vagas. A notação, a que chamamos tipo 2, diverge da notação até aqui utilizada (tipo 1) na quantidade de informação colocada no atributo TIPOREL: (i) a faceta da própria EM (CATORIGEM) que entra na relação, (ii) o tipo da relação (TIPOREL), (iii) o ID da EM relacionada (COREL) e (iv) a faceta da EM relacionada (CATALVO), separados por **.

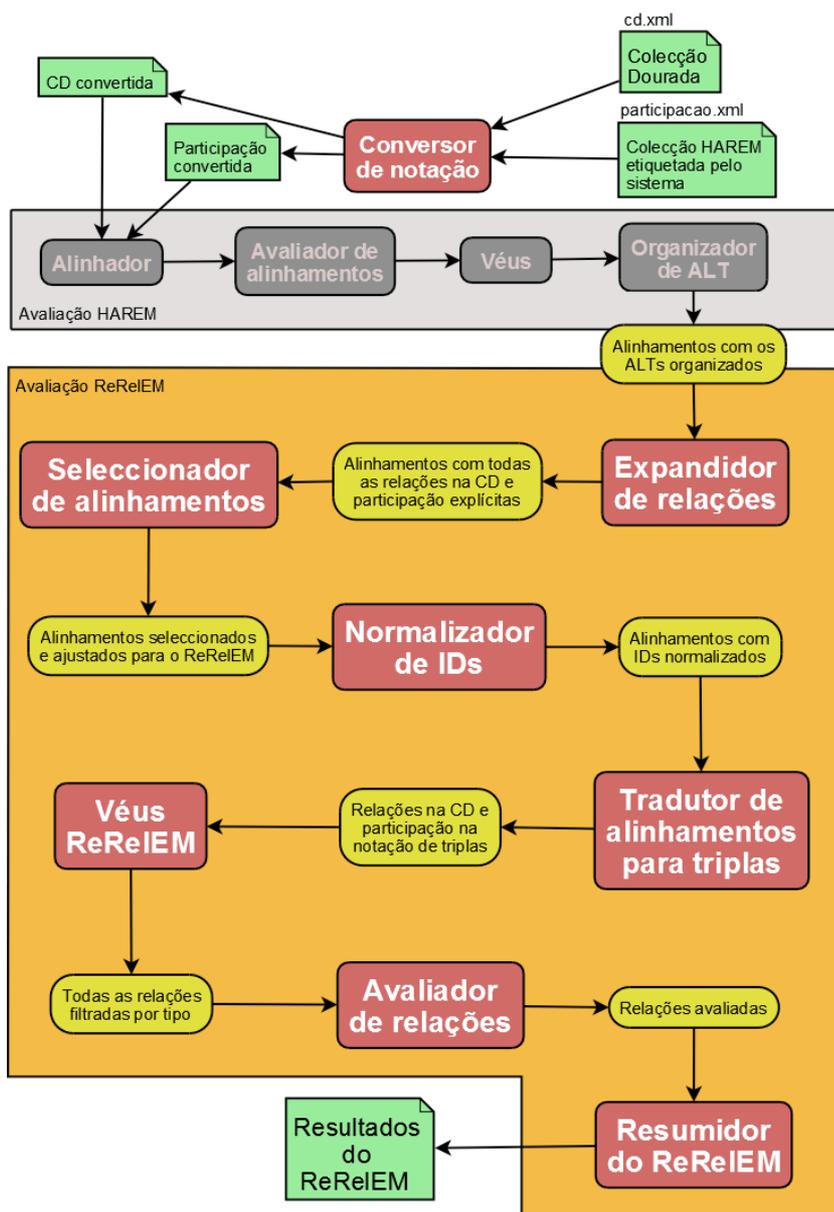


Figura 5.15: Arquitectura dos programas de avaliação para o ReReEM.

```

Tipo 1: <EM ID="X" CATEG="CATORIGEM1|CATORIGEM2" COREL="ID1_ID1_ID2" TIPOREL="re11_re12_re13">
Tipo 1 (com especificação de facetas relacionadas): <EM ID="X" CATEG="CATORIGEM1|CATORIGEM2"
COREL="ID1_ID2_ID2" TIPOREL="re11_re12_re13" FAC_ORIGEM="CATORIGEM1_CATORIGEM2_
CATORIGEM1" FAC_ALVO="CATALVO1_CATALVO1_CATALVO1">
Tipo 2: <EM ID="X" CATEG="CATORIGEM1|CATORIGEM2" TIPOREL="CATORIGEM1**re11**ID1**CATALVO1_
CATORIGEM2**re12**ID1**CATALVO1_CATORIGEM1**re13**ID2**CATALVO1">

```

Figura 5.16: Formatos da CD do ReRelEM

Notação de tipo 1:

```

<EM ID="X" CATEG="LOCAL">Europa</EM>
<EM ID="Y" CATEG="LOCAL|ORGANIZACAO">UE</EM>
<EM ID="Z" CATEG="LOCAL|ORGANIZACAO" COREL="X_Y_Y" TIPOREL="incluido_incluido_incluido" FAC_
_ORIGEM="LOCAL_LOCAL_ORGANIZACAO" FAC_ALVO="LOCAL_LOCAL_ORGANIZACAO">Portugal</EM>

```

Notação de tipo 2:

```

<EM ID="X" CATEG="LOCAL">Europa</EM>
<EM ID="Y" CATEG="LOCAL|ORGANIZACAO">UE</EM>
<EM ID="Z" CATEG="LOCAL|ORGANIZACAO" COREL="X_Y_Y" TIPOREL="LOCAL**incluido**X**LOCAL_LOCAL
**incluido**Y**LOCAL_ORGANIZACAO**incluido**Y**ORGANIZACAO">Portugal</EM>

```

Figura 5.17: Entidades anotadas com notações de tipo 1 e 2

Apesar da notação de tipo 2 facilitar a anotação e leitura das relações por humanos, o seu processamento coloca alguns problemas técnicos⁶. Foi por isso decidido que, embora mantida na CD do ReRelEM, esta notação seria transformada durante o processo de avaliação, a fim de manter a compatibilidade com os demais programas. Assim, o atributo TIPOREL volta a ter apenas o tipo da relação, e cada EM recebe dois novos atributos para representar as facetas nas relações (FAC_ORIGEM e FAC_ALVO). A essa notação chamamos tipo 1 com especificação de facetas, porque o preenchimento do atributo TIPOREL é igual ao tipo 1, e a diferença encontra-se nos dois atributos adicionais.

A figura 5.16 ilustra genericamente as notações de tipo 1 (com e sem facetas especificadas) e tipo 2.

Para cada relação que dá entrada nos programas é obrigatória uma entrada nos atributos FAC_ORIGEM e FAC_ALVO. Quando se trata de uma relação entre EM vagas e existe a informação das facetas relacionadas (no tipo 2), as facetas são obtidas através da informação no atributo TIPOREL. Quando essa informação não existe (EM simples ou EM numa participação onde essa informação não existe), considera-se que as categorias das EM relacionadas são também as facetas na relação. Na figura 5.17, ilustramos várias entidades anotadas tanto com o tipo 1 como 2.

⁶ Dificuldades relativas às várias informações que passam a estar contidas no atributo TIPOREL e a replicação, também no TIPOREL, do ID, que já se encontra no atributo COREL.

5.8.2 Expandidor de relações

O Expandidor de relações é um módulo chave no ReRelEM, utilizado para maximizar todas as relações implícitas no conjunto de relações anotadas na CD ou nas participações.

O conjunto de relações anotadas é aumentado por meio da explicitação de todas as relações inversas ou simétricas (se existirem) que não se encontram marcadas. O módulo testa ainda se a combinação de cada par de relações existente pode originar uma nova relação através das regras de expansão. A informação relativa aos tipos de relação e respectivas inversas encontra-se no ficheiro `harem.conf` de forma a poder ser facilmente parametrizada.

5.8.2.1 Expansão

As regras de expansão utilizadas neste primeiro ReRelEM resumem-se às seguintes quatro (já mencionadas no capítulo 4):

1. $A \text{ ident } B \wedge B \text{ ident } C \Rightarrow A \text{ ident } C$
2. $A \text{ inclui } B \wedge B \text{ inclui } C \Rightarrow A \text{ inclui } C$
3. $A \text{ inclui } B \wedge B \text{ sede_de } C \Rightarrow A \text{ sede_de } C$
4. $A \text{ ident } B \wedge B \text{ qualquer_relação } C \Rightarrow A \text{ qualquer_relação } C$

O Expandidor procura normalizar cada par de relações de forma a verificar se algum segue uma das regras anteriores. A normalização passa essencialmente pelos passos descritos a seguir:

1. verificar se as duas relações do par têm um argumento comum (B) e outro diferente;
2. transformar todas as relações que têm inversa no tipo convencional directo (`inclui`, `sede_de`, ...);⁷
3. colocar como primeiro elemento do par a relação que tiver B no segundo argumento, e como segundo elemento do par, a relação que tiver B no primeiro argumento;
4. se o par obtido seguir uma das regras de expansão, ele dá origem a uma nova relação que, se ainda não existir, será adicionada ao conjunto de relações existentes.

Se algum destes passos não for possível, o par não dará origem a uma nova relação. Na figura 5.18 estão alguns exemplos do procedimento de expansão.

É possível optar pela expansão apenas das relações do lado da CD (`-exptudo nao`), embora por omissão a expansão seja efectuada dos dois lados. Esta opção foi desenvolvida para verificar as consequências da expansão de eventuais relações erradas.

⁷ Como a relação de identidade é simétrica, esta transformação é realizada se necessário.

Par	Passo 1	Passo 2	Passo 3	Resultado
$B \text{ ident } C \wedge A \text{ ident } B$	Sim	$B \text{ ident } C \wedge A \text{ ident } B$	$A \text{ ident } B \wedge B \text{ ident } C$	$A \text{ ident } C$
$A \text{ incluído } B \wedge C \text{ incluído } A$	Sim	$B \text{ incluído } A \wedge C \text{ incluído } A$	Impossível	Nada
$A \text{ incluído } B \wedge A \text{ incluído } C$	Sim	$B \text{ incluído } A \wedge A \text{ incluído } C$	$B \text{ incluído } A \wedge A \text{ incluído } C$	$B \text{ incluído } C$
$A \text{ ocorre_em } B \wedge B \text{ incluído } C$	Sim	$B \text{ ocorre_em } A \wedge C \text{ incluído } B$	$C \text{ incluído } B \wedge B \text{ ocorre_em } A$	$C \text{ ocorre_em } A$
$A \text{ ocorre_em } B \wedge C \text{ incluído } D$	Não	Nada	Nada	Nada
$B \text{ ident } A \wedge B \text{ natural_de } C$	Sim	$B \text{ ident } A \wedge B \text{ natural_de } C$	$A \text{ ident } B \wedge B \text{ natural_de } C$	$A \text{ natural_de } C$

Figura 5.18: Exemplos de aplicação das regras de transitividade

5.8.2.2 Compatibilidade de facetas

Consideramos as facetas intervenientes em cada relação do lado da participação as próprias categorias das EM, sejam elas simples ou vagas.⁸ Definiu-se por isso que uma relação é igual a outra se tiver o mesmo tipo, os mesmos argumentos (em termos de EM) e facetas compatíveis, ou seja, facetas iguais ou facetas passíveis de unificação, considerando que uma categoria não existente (se por acaso o sistema apenas marcou EM e não especificou a categoria) é compatível com qualquer conjunto de categorias. Por exemplo:

- LOCAL é compatível com LOCAL;
- LOCAL|ORGANIZACAO é compatível com LOCAL;
- EM é compatível com LOCAL.

5.8.3 Seleccionador de alinhamentos

O Seleccionador de alinhamentos é aplicado ao conjunto de alinhamentos produzidos pelo Organizador de ALT, de onde vai seleccionar apenas os alinhamentos (ou parte) relativos a respostas correctas dadas pelos sistemas no HAREM clássico, de acordo com o descrito na secção 5.3.3.

5.8.4 Normalizador de identificadores (ID)

Para que seja possível comparar as relações do lado da CD com as relações do lado da participação é necessário normalizar os ID das EM de cada lado, utilizando um identificador comum. O Normalizador de identificadores vai substituir o valor dos atributos ID das EM do lado da participação pelo valor dos atributos ID das EM correspondentes na CD. As entradas dos atributos COREL, TIPOREL, FAC_ORIGEM e FAC_ALVO são também normalizadas: os valores dos ID em COREL são adaptados aos novos ID e as relações que envolvam pelo menos um ID ou um conjunto *ID+faceta* inexistente são removidas.

5.8.5 Tradutor de alinhamentos para triplas

Como o foco do ReReLEM é a identificação e avaliação de relações, para facilitar a depuração e o processamento pelos módulos seguintes, o módulo Tradutor de alinhamentos para triplas transforma os vários alinhamentos em listas de relações, representadas por triplas. Para cada documento são listadas as relações do lado da CD e as relações do lado da participação. As triplas têm o formato: *arg1 tiporel arg2*, onde *arg1* e *arg2* são os argumentos da relação e podem ser representados por um ID (de uma EM) ou por um ID seguido de uma categoria (faceta).

Ilustramos a conversão da notação de alinhamentos na notação de triplas exemplificando a entrada e a saída do Tradutor nas figuras 5.19 e 5.20, respectivamente.

Neste caso a entrada do módulo é um ficheiro de alinhamentos com ID normalizados. Como se verifica, as relações representadas nos alinhamentos através dos atributos COREL, TIPOREL, FAC_ORIGEM e FAC_ALVO são transformadas em duas listas: as relações do lado da CD e as relações do lado da participação.

⁸ Note-se que, nas participações, as relações nunca podiam estar definidas entre facetas, visto que isto foi algo que a organização marcou após a ocorrência da própria avaliação conjunta no sentido estrito.


```

DOC hub-66526
[CD]
hub-66526-556 LOCAL inclui hub-66526-568 LOCAL
hub-66526-556 LOCAL sede_de hub-66526-542 ORGANIZACAO
hub-66526-556 LOCAL inclui hub-66526-538 LOCAL
hub-66526-556 LOCAL inclui hub-66526-540 LOCAL
hub-66526-556 LOCAL inclui hub-66526-564 LOCAL
hub-66526-556 LOCAL inclui hub-66526-542 LOCAL
hub-66526-556 LOCAL sede_de hub-66526-564 ORGANIZACAO
hub-66526-542 LOCAL sede_de hub-66526-564 ORGANIZACAO
hub-66526-542 LOCAL incluído hub-66526-568 LOCAL
hub-66526-542 LOCAL sede_de hub-66526-542 ORGANIZACAO
hub-66526-542 LOCAL ident hub-66526-564 LOCAL
hub-66526-542 LOCAL incluído hub-66526-556 LOCAL
hub-66526-542 ORGANIZACAO ident hub-66526-564 ORGANIZACAO
hub-66526-542 ORGANIZACAO ocorre_em hub-66526-556 LOCAL
hub-66526-542 ORGANIZACAO ocorre_em hub-66526-568 LOCAL
hub-66526-542 ORGANIZACAO ocorre_em hub-66526-564 LOCAL
hub-66526-542 ORGANIZACAO ocorre_em hub-66526-542 LOCAL
hub-66526-25 PESSOA ident hub-66526-560 PESSOA
hub-66526-564 LOCAL incluído hub-66526-568 LOCAL
hub-66526-564 LOCAL sede_de hub-66526-542 ORGANIZACAO
hub-66526-564 LOCAL ident hub-66526-542 LOCAL
hub-66526-564 LOCAL incluído hub-66526-556 LOCAL
hub-66526-564 LOCAL sede_de hub-66526-564 ORGANIZACAO
hub-66526-560 PESSOA ident hub-66526-25 PESSOA
hub-66526-540 LOCAL incluído hub-66526-556 LOCAL
hub-66526-568 LOCAL inclui hub-66526-542 LOCAL
hub-66526-568 LOCAL inclui hub-66526-564 LOCAL
hub-66526-568 LOCAL incluído hub-66526-556 LOCAL
hub-66526-568 LOCAL sede_de hub-66526-542 ORGANIZACAO
hub-66526-568 LOCAL sede_de hub-66526-564 ORGANIZACAO
hub-66526-564 ORGANIZACAO ocorre_em hub-66526-556 LOCAL
hub-66526-564 ORGANIZACAO ocorre_em hub-66526-568 LOCAL
hub-66526-564 ORGANIZACAO ocorre_em hub-66526-564 LOCAL
hub-66526-564 ORGANIZACAO ident hub-66526-542 ORGANIZACAO
hub-66526-564 ORGANIZACAO ocorre_em hub-66526-542 LOCAL
hub-66526-538 LOCAL incluído hub-66526-556 LOCAL
[Part]
hub-66526-25 PESSOA ident hub-66526-560 PESSOA
hub-66526-564 LOCAL|ORGANIZACAO ident hub-66526-542 LOCAL|ORGANIZACAO
hub-66526-560 PESSOA ident hub-66526-25 PESSOA
hub-66526-540 LOCAL ocorre_em hub-66526-538 LOCAL
hub-66526-542 LOCAL|ORGANIZACAO ident hub-66526-564 LOCAL|ORGANIZACAO
hub-66526-538 LOCAL sede_de hub-66526-540 LOCAL
EOD

```

Figura 5.20: Saída do Tradutor de alinhamentos para triplas.

5.8.6 Véus para o ReReLEM

Este módulo é semelhante ao Véus do HAREM clássico, mas aplicado aos tipos de relação: tem como entrada um filtro constituído por um conjunto de tipos de relação, separadas pelo carácter `'`. As relações de tipos que não se encontrem no filtro são removidas tanto do lado da CD como do lado da participação. A saída deste módulo terá no cabeçalho não apenas a descrição do cenário selectivo do HAREM em que é feita a avaliação, mas também a descrição do cenário do ReReLEM, sendo, de resto, em tudo semelhante à saída do `Alinhamentos para triplas`.

5.8.7 Avaliador de relações

O `Avaliador de relações` compara as relações do lado da participação com as relações do lado da CD.

Antes da comparação ser feita, todas as relações da CD que especializam/explicitam a relação `outra` são convertidas outra vez nesta mais geral (ver secção 4.1.4), de acordo com a definição original da tarefa.

Em seguida, são consideradas correctas todas as relações que estão na participação e para as quais existe na CD uma relação compatível. Considera-se que duas relações são compatíveis se os seus argumentos forem iguais, as suas facetas forem compatíveis (ver secção 5.8.2.2) e se o seu tipo for igual.

O resultado da aplicação do `Avaliador` é um conjunto de relações correctas, espúrias e em falta, ao qual é anexado o número total de relações na CD (`Rels_CD`) e o número total de relações na participação (`Rels_Part`).

A figura 5.21 mostra um excerto da saída do avaliador de relações, com todas as relações de determinado documento avaliadas.

5.8.7.1 Resumidor das classificações do ReReLEM

O `Resumidor das classificações do ReReLEM` processa a saída do `Avaliador de relações` agregando os resultados para obter os resultados globais da avaliação do ReReLEM na participação. Na figura 5.22 encontra-se a título de exemplo uma saída deste módulo, onde se podem ver as métricas e medidas calculadas.

5.8.8 Gerador de resultados do ReReLEM

O `Gerador de resultados do ReReLEM` tem uma função equivalente à do `Gerador de resultados do HAREM` (secção 5.6.9), agrupando todos os resultados do ReReLEM num ficheiro em HTML com os sistemas participantes ordenados por medida F.

5.8.9 Visualizador de relações

O `Visualizador de relações` é um programa que desenha grafos de relações a partir de ficheiros anotados segundo as directivas do ReReLEM. Tem como entrada (se nada for especificado) um ficheiro de alinhamentos, como por exemplo a saída do `Organizador de ALT`, do `Expandidor de relações`, do `Seleccionador de alinhamentos` ou do `Normalizador de ID`. É também possível ter uma entrada sob a forma de ficheiro de relações representadas através de triplas (utilizando a opção `-entrada triplas`).

```

DOC hub-94570
[Rels_CD(33) Rels_Part(10)]
hub-94570-125 LOCAL ident hub-94570-128 LOCAL :: Aval(Relação Correcta)
hub-94570-128 LOCAL ident hub-94570-125 LOCAL :: Aval(Relação Correcta)
hub-94570-114 PESSOA ident hub-94570-115 PESSOA :: Aval(Relação Correcta)
hub-94570-115 PESSOA ident hub-94570-114 PESSOA :: Aval(Relação Correcta)
hub-94570-112 PESSOA incluido hub-94570-114 PESSOA :: Aval(Espuria)
hub-94570-112 PESSOA incluido hub-94570-115 PESSOA :: Aval(Espuria)
hub-94570-118 ORGANIZACAO ocorre_em hub-94570-131 LOCAL :: Aval(Espuria)
hub-94570-131 LOCAL sede_de hub-94570-118 ORGANIZACAO :: Aval(Espuria)
hub-94570-114 PESSOA inclui hub-94570-112 PESSOA :: Aval(Espuria)
hub-94570-115 PESSOA inclui hub-94570-112 PESSOA :: Aval(Espuria)
hub-94570-110 TEMPO incluido hub-94570-117 TEMPO :: Aval(Em Falta)
hub-94570-112 PESSOA ident hub-94570-114 PESSOA :: Aval(Em Falta)
hub-94570-112 PESSOA ident hub-94570-115 PESSOA :: Aval(Em Falta)
hub-94570-112 PESSOA outra hub-94570-113 PESSOA :: Aval(Em Falta)
hub-94570-112 PESSOA outra hub-94570-118 ORGANIZACAO :: Aval(Em Falta)
hub-94570-118 ORGANIZACAO outra hub-94570-112 PESSOA :: Aval(Em Falta)
hub-94570-118 ORGANIZACAO outra hub-94570-113 PESSOA :: Aval(Em Falta)
hub-94570-118 ORGANIZACAO outra hub-94570-114 PESSOA :: Aval(Em Falta)
hub-94570-118 ORGANIZACAO outra hub-94570-115 PESSOA :: Aval(Em Falta)
hub-94570-125 LOCAL outra hub-94570-116 PESSOA :: Aval(Em Falta)
hub-94570-113 PESSOA outra hub-94570-114 PESSOA :: Aval(Em Falta)
hub-94570-113 PESSOA outra hub-94570-115 PESSOA :: Aval(Em Falta)
hub-94570-113 PESSOA outra hub-94570-112 PESSOA :: Aval(Em Falta)
hub-94570-113 PESSOA outra hub-94570-118 ORGANIZACAO :: Aval(Em Falta)
hub-94570-128 LOCAL outra hub-94570-116 PESSOA :: Aval(Em Falta)
hub-94570-116 PESSOA outra hub-94570-128 LOCAL :: Aval(Em Falta)
hub-94570-116 PESSOA outra hub-94570-125 LOCAL :: Aval(Em Falta)
hub-94570-117 TEMPO outra hub-94570-114 PESSOA :: Aval(Em Falta)
hub-94570-117 TEMPO inclui hub-94570-111 TEMPO :: Aval(Em Falta)
hub-94570-117 TEMPO inclui hub-94570-110 TEMPO :: Aval(Em Falta)
hub-94570-117 TEMPO outra hub-94570-115 PESSOA :: Aval(Em Falta)
hub-94570-117 TEMPO outra hub-94570-112 PESSOA :: Aval(Em Falta)
hub-94570-114 PESSOA outra hub-94570-118 ORGANIZACAO :: Aval(Em Falta)
hub-94570-114 PESSOA ident hub-94570-115 PESSOA :: Aval(Em Falta)
hub-94570-114 PESSOA outra hub-94570-113 PESSOA :: Aval(Em Falta)
hub-94570-111 TEMPO incluido hub-94570-117 TEMPO :: Aval(Em Falta)
hub-94570-115 PESSOA outra hub-94570-118 ORGANIZACAO :: Aval(Em Falta)
hub-94570-115 PESSOA outra hub-94570-113 PESSOA :: Aval(Em Falta)
hub-94570-115 PESSOA ident hub-94570-114 PESSOA :: Aval(Em Falta)
EOD

```

Figura 5.21: Fragmento da saída do Avaliador de relações

```

Total na CD: 1626
Total identificadas: 722
Total correctamente identificadas: 409
Espúrios: 313
Em Falta: 1217
Precisão: 0.5664819944598338
Abrangência: 0.25153751537515373
Medida F: 0.34838160136286195

```

Figura 5.22: Fragmento da saída do Resumidor das classificações do ReReLEM

Como a entrada pode consistir num conjunto de relações ou alinhamentos para vários documentos, é também possível escolher a origem do grafo a visualizar, indicando o ID do documento e o lado (CD ou participação).

No grafo, as relações podem ser identificadas pela cor dos arcos: *ident* (verde), *inclui* (vermelho), *incluído* (violeta), *ocorre_em* (azul), *sede_de* (azul claro) e *outra* (preto).

As figuras 5.23 e 5.24 correspondem a um grafo do mesmo documento, antes da expansão (5.23) e depois de ser expandido (5.24).

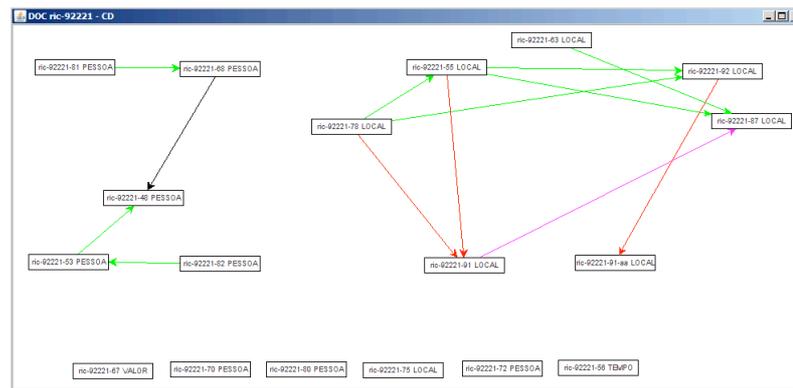


Figura 5.23: Grafo obtido através do Visualizador de relações sobre um conjunto de relações antes da expansão (saída do Normalizador de ID).

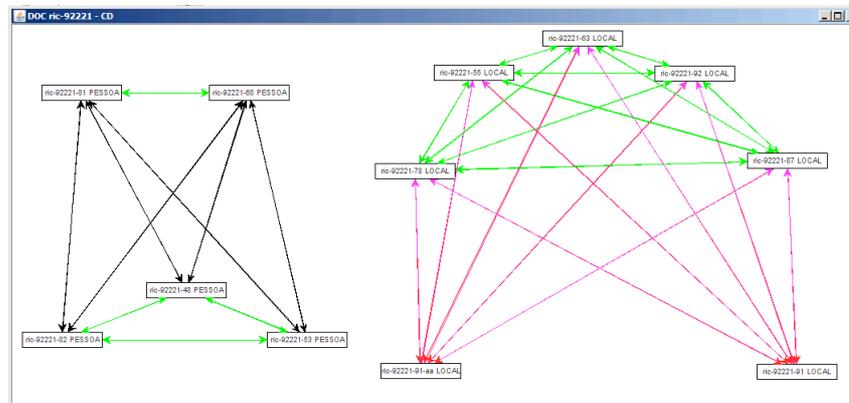


Figura 5.24: Grafo obtido através do Visualizador de relações sobre um conjunto de relações expandidas (saída do Expansidor de relações).

5.9 Observações finais

Todos os programas descritos podem ser invocados através de um serviço na rede, o SAHARA (descrito no apêndice G). Além disso, e como é norma da Linguateca, o có-

digo foi disponibilizado ao público. Todos os programas se encontram assim incluídos no pacote de recursos do Segundo HAREM, a LÂMPADA (disponível de <http://www.linguateca.pt/HAREM/PacoteRecursosSegundoHAREM.zip>), sob licença BSD.

Pensamos com esta postura aumentar o impacto e a utilidade da nossa actividade de organização de avaliações conjuntas, e contribuir para a melhoria do parque informático em termos de processamento do português.

Agradecimentos

Agradecemos ao Nuno Cardoso e ao Alberto Simões as valiosas sugestões de enriquecimento e clarificação do presente capítulo.