

Capítulo 18

Avaliação no HAREM: métodos e medidas

Diana Santos, Nuno Cardoso e Nuno Seco

Este capítulo foi previamente publicado como Relatório Técnico DI/FCUL TR-06-17, Departamento de Informática, Faculdade de Ciências da Universidade de Lisboa.

Diana Santos e Nuno Cardoso, editores, *Reconhecimento de entidades mencionadas em português: Documentação e actas do HAREM, a primeira avaliação conjunta na área*, Capítulo 18, p. 245–282, 2007.

As directivas de avaliação descritas neste relatório técnico representam o conjunto de pontuações, regras, medidas e métricas usadas para medir e comparar as saídas dos sistemas de REM em relação às colecções douradas. O *software* de avaliação do HAREM, descrito ao detalhe no capítulo 19, implementa as directivas aqui expostas.

18.1 Terminologia

18.1.1 Pontuações

As *pontuações* são os valores atribuídos a cada EM marcada pelo sistema, após uma comparação com a respectiva marcação na CD. Cada tarefa possui as suas próprias pontuações, que são calculadas segundo um conjunto de *regras* que serão descritas e ilustradas com exemplos nas respectivas secções.

Um exemplo simples de pontuação: se um sistema identificar uma determinada EM tal como está na CD, obtém a pontuação *correcto* para a tarefa de identificação. A pontuação *correcto*, segundo as regras para a tarefa de identificação, corresponde a um valor igual a 1.

18.1.2 Medidas

As *medidas* representam formas de combinação das várias pontuações obtidas em cada tarefa. Assim, é possível representar diferentes componentes da avaliação, para cada saída.

As medidas são implementadas na avaliação das tarefas de classificação morfológica e de semântica, onde existe mais do que um parâmetro pontuável (no caso da morfologia, o género e o número; no caso da semântica, a categoria e o tipo).

Um exemplo de medidas: na tarefa de classificação semântica, uma EM é avaliada segundo a sua categoria e tipo. Assim, são geradas duas pontuações, uma relacionada com a categoria, e outra com o tipo. A combinação destas pontuações num único valor depende da *medida* usada.

18.1.3 Métricas

As *métricas* são formas de representar o desempenho dos sistemas em valores numéricos, de acordo com a marcação que fez para um dado grupo de EM.

Precisão: a precisão afere a “qualidade” da resposta do sistema, ao calcular a proporção de respostas correctas em relação a todas as respostas realizadas por este.

Abrangência: a abrangência afere a “quantidade” da resposta do sistema, ao calcular a proporção de respostas correctas em relação ao universo de possíveis respostas (no caso presente, as EM contidas na colecção dourada).

Medida F: A medida F combina as métricas de precisão e de abrangência para cada tarefa, de acordo com a seguinte fórmula:

$$\text{Medida F} = \frac{2 \times \text{precisão} \times \text{abrangência}}{\text{precisão} + \text{abrangência}}$$

Esta métrica é igual para todas as tarefas de avaliação.

Sobre-geração: a sobre-geração afere o excesso de resultados que um sistema produz, ou seja, calcula quantas vezes produz resultados espúrios.

Sub-geração: a sub-geração afere a quantidade de resultados que um sistema se esqueceu em analisar, ou seja, calcula quantas vezes produz resultados em falta, dada a solução conhecida (a CD).

Erro Combinado: o erro combinado reúne as métricas de sobre-geração e de sub-geração numa única métrica, de acordo com a seguinte fórmula:

$$\text{Erro combinado} = \frac{\sum \text{pontuações em falta} + \sum \text{pontuações espúrio} + \sum \text{factor de erro}}{\sum \text{Pontuação máx. sistema} \cup \text{Pontuação máx. CD}}$$

O factor de erro é calculado pela equação 18.2, apresentada mais à frente.

18.1.4 Cenários de avaliação

Os sistemas de REM são desenvolvidos para diferentes propósitos. Como tal, as directivas de avaliação prevêm a realização de avaliações segundo *cenários*, de forma a ajustar a avaliação às características de cada sistema de REM. O módulo de *software* responsável pela criação de cenários é o Véus, que se encontra detalhado no capítulo 19.

A avaliação do HAREM realizou-se segundo dois eixos de cenários:

Cenário absoluto–relativo: O cenário absoluto avalia o desempenho do sistema em relação à totalidade das EM na CD para a tarefa de REM completa, ou seja, a identificação e a classificação de EM. O cenário relativo, por seu lado, restringe a avaliação às EM pontuadas como correcto ou parcialmente correcto na tarefa de identificação. Este cenário permite avaliar o desempenho do sistema apenas na tarefa de classificação (semântica ou morfológica), independentemente do desempenho na tarefa de identificação.

Cenário total–selectivo O cenário total abrange todas as categorias de EM da CD, avaliando a tarefa de classificação (morfológica ou semântica) em relação à tarefa tal como foi proposta pelo HAREM. No cenário selectivo, o participante escolhe previamente um sub-conjunto de categorias e de tipos da categorização HAREM que o seu sistema

consegue processar. Assim, a tarefa da classificação (morfológica ou semântica) é avaliada segundo esse sub-conjunto de categorias e de tipos.

Tarefa de identificação

A tarefa de identificação é avaliada segundo o eixo de cenário total–selectivo:

Cenário de identificação total: considera para efeitos de pontuação todas as etiquetas na CD.

Cenário de identificação selectivo: considera apenas para efeitos de pontuação o leque de categorias semânticas que o sistema participante se propõe explicitamente identificar.

Tarefas de classificação

As tarefas de classificação (morfológica e semântica) são avaliadas segundo os dois eixos de cenários:

Total: considera todas as EM existentes na CD.

Absoluto: considera todas as EM, incluindo as que não foram identificadas com pontuação correcta ou parcialmente correcta.

Relativo: considera apenas as EM identificadas com pontuação correcta ou parcialmente correcta.

Selectivo: considera apenas as EM na CD de categorias/tipos que o participante se propôs classificar.

Absoluto: considera todas as EM, incluindo as que não foram identificadas com pontuação correcta ou parcialmente correcta.

Relativo: considera apenas as EM identificadas com pontuação correcta ou parcialmente correcta. correctas.

18.2 Tarefa de identificação

A avaliação da tarefa de identificação tem por objectivo medir a eficiência dos sistemas em delimitar correctamente os *termos* que compõem as EM na colecção, comparativamente com a CD).

Um termo é definido no HAREM como sendo qualquer sequência de letras (e somente letras) ou dígitos individuais. As preposições e conjunções são contabilizadas para efeitos de pontuação, mas não são considerados para efeitos de alinhamento. No capítulo 19, secção 19.2.3, está disponível uma lista das palavras que o AlinhEM, o módulo de *software* que realiza os alinhamentos, ignora, assim como as regras de atomização.

18.2.1 Pontuações

A avaliação do HAREM atribui a seguinte pontuação para a tarefa de identificação:

Correcto: quando o termo inicial e o termo final da EM são iguais na saída e na CD, e o número de termos da EM é o mesmo nas duas listas.

Parcialmente Correcto (por defeito): quando pelo menos um termo da saída do sistema corresponde a um termo de uma EM na CD, e o número total de termos da EM do sistema é menor do que o número de termos da respectiva EM da CD.

Parcialmente Correcto (por excesso): quando pelo menos um termo da saída do sistema corresponde a um termo de uma EM na CD, e o número total de termos da EM do sistema é maior do que o número de termos da respectiva EM da CD.

Em Falta: quando a saída do sistema falha em delimitar correctamente qualquer termo de uma EM da CD.

Espúrio: quando a saída do sistema delimita uma alegada EM que não consta na CD.

Às EM pontuadas como *correcto* é atribuído um valor igual a 1. As EM pontuadas como *parcialmente correcto* é atribuído o valor calculado pela equação 18.1:

$$p = 0,5 \frac{n_c}{n_d} \quad (18.1)$$

Onde:

n_c representa o número de termos comuns entre a EM do sistema e a EM da CD, ou seja, a cardinalidade da intersecção dos termos.

n_d representa o número de termos distintos entre a EM do sistema e a EM da CD, ou seja, a cardinalidade da reunião dos termos.

O *factor de erro*, usado no cálculo da métrica *Erro Combinado*, é dado pela equação 18.2:

$$p = 1 - 0,5 \frac{n_c}{n_d} \quad (18.2)$$

18.2.2 Métricas

Para a tarefa de identificação, as métricas são calculadas da seguinte forma:

Precisão

Na tarefa de identificação, a precisão calcula o teor de EM correctas e parcialmente correctas em todas as EM identificadas pelo sistema. Os valores para as EM pontuadas como parcialmente correctas são calculados pela equação 18.1.

$$\text{Precisão}_{\text{identificação}} = (\sum \text{EM correctas} + \sum \text{EM parcialmente correctas}) / (\sum \text{EM identificadas pelo sistema})$$

Abrangência

Na tarefa de identificação, a abrangência calcula o teor de EM contidas na CD que o sistema conseguiu identificar. Os valores para as EM pontuadas como parcialmente correctas são calculados pela equação 18.1.

$$\text{Abrangência}_{\text{identificação}} = (\sum \text{EM correctas} + \sum \text{EM parcialmente correctas}) / (\sum \text{EM na CD})$$

Sobre-geração

Na tarefa de identificação, a sobre-geração calcula o teor de EM que foram identificadas pelo sistema, mas que não existem na CD.

$$\text{Sobre-geração}_{\text{identificação}} = (\sum \text{EM espúrias} / \sum \text{EM identificadas pelo sistema})$$

Sub-geração

Na tarefa de identificação, a sub-geração calcula o teor de EM que existem na colecção dourada, mas que não foram identificadas pelo sistema.

$$\text{Sub-geração}_{\text{identificação}} = (\sum \text{EM em falta} / \sum \text{EM na CD})$$

18.2.3 Exemplo detalhado de atribuição de pontuação

Apresentemos uma frase hipotética da colecção dourada:

Terminou ontem no <LOCAL TIPO="ALARGADO"> Laboratório Nacional de Engenharia Civil </LOCAL>, em <LOCAL TIPO="ADMINISTRATIVO"> Lisboa </LOCAL>, o <ACONTECIMENTO TIPO="EVENTO"> Encontro de Reflexão </ACONTECIMENTO> sobre a concretização do <ABSTRACCAO TIPO="PLANO"> Plano Hidrológico </ABSTRACCAO> espanhol.

Imaginemos a seguinte saída do sistema que pretendemos avaliar:

```
<PESSOA TIPO="INDIVIDUAL">Terminou</PESSOA> ontem no
<LOCAL TIPO="ALARGADO">Laboratório Nacional</LOCAL>
de <ABSTRACCAO TIPO="DISCIPLINA">Engenharia
Civil</ABSTRACCAO>, em <LOCAL TIPO="ADMINISTRATIVO">Lisboa</LOCAL>,
o Encontro de Reflexão sobre a concretização do <ABSTRACCAO
TIPO="PLANO">Plano Hidrológico espanhol</ABSTRACCAO>.
```

A Tabela 18.1 apresenta a pontuação pormenorizada, caso a caso, e na Tabela 18.2 os valores das métricas para a tarefa de identificação. A Tabela 18.3 apresenta 7 casos particulares com uma anotação (hipotética) feita por um sistema e na CD, e a Tabela 18.4 ilustra as regras de pontuação para esses casos, para a tarefa de identificação.

18.2.4 Identificações alternativas

No caso de considerarmos que há mais do que uma delimitação correcta na tarefa em questão, levando à identificação de uma ou mais EM alternativas, foi usada a etiqueta <ALT> para assinalar as várias opções na CD. Como tal, o avaliador do HAREM irá comparar a CD com a saída do sistema e optar pela melhor alternativa. A escolha é feita segundo o seguinte algoritmo:

- 1º – Melhor medida F para cada caso.
- 2º – Menor valor de erro combinado.
- 3º – Maior número de alinhamentos.

Para auxiliar na selecção da opção <ALT> nos casos mais difíceis, tais como alternativas sem EM, os programas de selecção de <ALT> para as tarefas do HAREM (*ALTinaID*, *ALTinaSEM* e *ALTinaMOR*, ver capítulo 19) introduzem no cálculo um alinhamento correcto em cada alternativa considerada. Tal introdução não prejudica a selecção, e evita que alternativas sem EM tenham uma medida F não definida (ou seja, zero no numerador e no denominador).

No Capítulo 19 explica-se em detalhe este processo do *ALTinaID*, e as Tabelas 18.5 a 18.9 ilustram como é feito esse cálculo, para o seguinte exemplo com três alternativas:

```
<ALT> <EM> Governo PSD de Cavaco Silva </EM>
<EM> Governo PSD </EM> de <EM> Cavaco Silva </EM>
Governo PSD de Cavaco Silva </ALT>
```

ALT1: Governo PSD de Cavaco Silva

Caso	Colecção dourada	Saída do sistema	Pontuação
1	-	Terminou	0 (Espúrio)
2	Laboratório Nacional de Engenharia Civil	Laboratório Nacional	$0,5 \times \frac{2}{5} = 0,2$ (Parcialmente Correcto por Defeito)
3	Laboratório Nacional de Engenharia Civil	Engenharia Civil	$0,5 \times \frac{2}{5} = 0,2$ (Parcialmente Correcto por Defeito)
4	Lisboa	Lisboa	1 (Correcto)
5	Encontro de Reflexão	-	0 (Em Falta)
6	Plano Hidrológico	Plano Hidrológico espanhol	$0,5 \times \frac{2}{3} = 0,333$ (Parcialmente Correcto Por Excesso)

Tabela 18.1: Pontuação da tarefa de identificação, para o exemplo dado.

Métrica	Valor
Precisão	$\frac{1+0,2+0,2+0,333}{5} = 34,7\%$
Abrangência	$\frac{1+0,2+0,2+0,333}{4} = 43,3\%$
Medida F	$\frac{2 \times 0,347 \times 0,433}{0,347+0,433} = 0,385$
Sobre-geração	$\frac{1}{5} = 20\%$
Sub-geração	$\frac{1}{4} = 25\%$
Erro Combinado	$\frac{(1-0,2)+(1-0,2)+(1-0,333)+1+1}{6} = 71,1\%$

Tabela 18.2: Métricas da tarefa de identificação, para o exemplo dado.

Caso	Sistema participante	Colecção dourada
1	o novo presidente do CNPq, Evando Mirra	o novo presidente do CNPq, Evando Mirra
2	a partir de 1991	a partir de 1991
3	Graduou-se em Engenharia Mecânica e Elétrica	Graduou-se em Engenharia Mecânica e Elétrica
4	Rua 13 de Maio, 733 - Bela Vista - (11) 3262 3256	Rua 13 de Maio, 733 - Bela Vista - (11) 3262 3256
5	Senhores Comandantes das F-FDTL e da PNTL	Senhores Comandantes das F-FDTL e da PNTL
6	secretário-geral do Partido Revolucionário Institucional	secretário-geral do Partido Revolucionário Institucional
7	Estúdio da Oficina Cultural Oswald de Andrade São Paulo, 21 de novembro de 1994	Estúdio da Oficina Cultural Oswald de Andrade São Paulo, 21 de novembro de 1994

Tabela 18.3: Lista de exemplos para ilustração da pontuação da tarefa de identificação.

Caso	Etiquetas	Pontuação	Termos	Total
1	Saída: presidente do CNPq, Evando CD: CNPq	$0,5 \times \frac{1}{4}$	n_c : CNPq n_d : presidente, do, CNPq, Evando	0,225
	Saída: presidente do CNPq, Evando CD: Evando Mirra	$0,5 \times \frac{1}{5}$	n_c : Evando n_d : presidente, do, CNPq, Evando, Mirra	
2	Saída: 991 CD: 1991	$0,5 \times \frac{3}{4}$	n_c : 9, 9, 1 n_d : 1, 9, 9, 1	0,375
3	Saída: Engenharia Mecânica CD: Engenharia Mecânica e Eléctrica	$0,5 \times \frac{2}{4}$	n_c : Engenharia, Mecânica n_d : Engenharia, Mecânica, e, Eléctrica	0,375
	Saída: Eléctrica CD: Engenharia Mecânica e Eléctrica	$0,5 \times \frac{1}{4}$	n_c : Eléctrica n_d : Engenharia, Mecânica, e, Eléctrica	
4	Saída: Rua CD: Rua 13 de Maio, 733 - Bela Vista	$0,5 \times \frac{1}{10}$	n_c : Rua n_d : Rua, 1, 3, de, Maio, 7, 3, 3, Bela, Vista	1,35
	Saída: 13 de Maio CD: Rua 13 de Maio, 733 - Bela Vista	$0,5 \times \frac{4}{10}$	n_c : 1, 3, de, Maio n_d : Rua, 1, 3, de, Maio, 7, 3, 3, Bela, Vista	
	Saída: Bela Vista CD: Rua 13 de Maio, 733 - Bela Vista	$0,5 \times \frac{2}{10}$	n_c : Bela, Vista n_d : Rua, 1, 3, de, Maio, 7, 3, 3, Bela, Vista	
	Saída: (11) 3262 3256 CD: (11) 3262 3256	1		
5	Saída: Senhores Comandantes das F- CD: Senhores Comandantes das F-FDTL e da PNTL	$0,5 \times \frac{4}{6}$	n_c : Senhores, Comandantes, das, F n_d : Senhores, Comandantes, das, F-, FDTL, PNTL	0,5
	Saída: FDTL CD: Senhores Comandantes das F-FDTL e da PNTL	$0,5 \times \frac{1}{6}$	n_c : FDTL n_d : Senhores, Comandantes, das, F-, FDTL, PNTL	
	Saída: PNTL CD: Senhores Comandantes das F-FDTL e da PNTL	$0,5 \times \frac{1}{6}$	n_c : PNTL n_d : Senhores, Comandantes, das, F-, FDTL, PNTL	
6	Saída: Partido Revolucionário Institucional CD: secretário-geral do Partido Revolucionário Institucional	$0,5 \times \frac{3}{6}$	n_c : Partido, Revolucionário, Institucional n_d : secretário, geral, do, Partido, Revolucionário, Institucional	0,25
7	Saída: Oficina Cultural Oswald de Andrade CD: Estúdio da Oficina Cultural Oswald de Andrade	$0,5 \times \frac{5}{6}$	n_c : Oficina, Cultural, Oswald, de, Andrade n_d : Estúdio, Oficina, Cultural, Oswald, de, Andrade	1,11(1)
	Saída: São Paulo, 21 CD: São Paulo	$0,5 \times \frac{2}{4}$	n_c : São, Paulo n_d : São, Paulo, 2, 1	
	Saída: São Paulo, 21 CD: 21 de novembro de 1994	$0,5 \times \frac{2}{9}$	n_c : 2, 1 n_d : 2, 1, de, Novembro, de, 1, 9, 9, 4	
	Saída: novembro de 1994 CD: 21 de novembro de 1994	$0,5 \times \frac{6}{9}$	n_c : Novembro, de, 1, 9, 9, 4 n_d : 2, 1, de, Novembro, de, 1, 9, 9, 4	

Tabela 18.4: Pontuação na tarefa de identificação, para os exemplos da tabela 18.3.

ALT2: Governo PSD de Cavaco Silva

ALT3: Governo PSD de Cavaco Silva

O avaliador irá escolher a alternativa que produz melhores resultados. A Tabela 18.5 apresenta vários exemplos de saídas de sistema (as células a negrito indicam a alternativa escolhida) e, para cada caso, a pontuação individual. Os valores da medida F e do erro combinado são calculados nas Tabelas 18.6 a 18.9, que se referem respectivamente à precisão, à abrangência, à medida F e ao erro combinado, escolhendo para cada caso qual das alternativas <ALT> será seleccionada. Como mencionado antes, as medidas nas Tabelas 18.6 a 18.9 são calculadas adicionando para cada alternativa um elemento correcto.

Caso	Sada do sistema	ALT1	ALT2	ALT3
1	Governo PSD de Cavaco Silva	1 Correcto Medida F: 1 Erro Combinado: 0%	2 Parc. Correcto Medida F: 0,56 Erro Combinado: 53,3%	1 Espúrio Medida F: 0,67 Erro Combinado: 50,0%
2	Governo PSD de Cavaco Silva	1 Em Falta Medida F: 0,67 Erro Combinado: 50,0%	2 Em Falta Medida F: 0,5 Erro Combinado: 66,7%	Sem pontuação Medida F: 1 Erro Combinado: 0%
3	Governo PSD de Cavaco Silva	1 Parc.Cor. por Def. Medida F: 0,7 Erro Combinado: 30%	2 Parc.Cor. por Exc. Medida F: 0,54 Erro Combinado: 55%	1 Espúrio Medida F: 0,67 Erro Combinado: 50%
4	Governo PSD de Cavaco Silva	2 Parc. Correcto Medida F: 0,48 Erro Combinado: 60%	2 Parc.Cor.+1 Em Falta Medida F: 0,5 Erro Combinado: 62,5%	2 Espúrio Medida F: 0,5 Erro Combinado: 66,7%
5	Governo PSD de Cavaco Silva	2 Parc. Correcto Medida F: 0,48 Erro Combinado: 60%	2 Parc. Correcto Medida F: 0,5 Erro Combinado: 50%	2 Espúrio Medida F: 0,48 Erro Combinado: 66,7%
6	Governo PSD de Cavaco Silva	1 Parc. Correcto Medida F: 0,6 Erro Combinado: 40%	1 Correcto, 1EmFalta Medida F: 0,8 Erro Combinado: 33,3%	1 Espúrio Medida F: 0,67 Erro Combinado: 50%
7	Governo PSD de Cavaco Silva	1 Parc. Correcto Medida F: 0,55 Erro Combinado: 45%	1 Parc. Cor., 1 Em Falta Medida F: 0,5 Erro Combinado: 58,3%	1 Espúrio Medida F: 0,67 Erro Combinado: 50%
8	Governo PSD de Cavaco Silva	1 Parc. Correcto Medida F: 0,651 Erro Combinado: 45%	2 Parc. Correcto Medida F: 0,5 Erro Combinado: 58,3%	1 Espúrio Medida F: 0,67 Erro Combinado: 50%

Tabela 18.5: Exemplos de selecção de alternativa na tarefa de identificação.

Caso	Precisão		
	ALT1	ALT2	ALT3
1	$(1+1)/(1+1)=100\%$	$(0,4+1)/(1+1)=70\%$	$(0+1)/(1+1)=50\%$
2	$(0+1)/(0+1)=100\%$	$(0+1)/(0+1)=100\%$	$(0+1)/(0+1)=100\%$
3	$(0,4+1)/(1+1)=70\%$	$(0,35+1)/(1+1)=67,5\%$	$(0+1)/(1+1)=50\%$
4	$(0,2+1)/(2+1)=40\%$	$(0,5+1)/(2+1)=50\%$	$(0+1)/(2+1)=33,3\%$
5	$(0,2+1)/(2+1)=40\%$	$(0,5+1)/(2+1)=50\%$	$(0+1)/(2+1)=33,3\%$
6	$(0,2+1)/(1+1)=60\%$	$(1+1)/(1+1)=100\%$	$(0+1)/(1+1)=50\%$
7	$(0,1+1)/(1+1)=55\%$	$(0,25+1)/(1+1)=62,5\%$	$(0+1)/(1+1)=50\%$
8	$(0,3+1)/(1+1)=65\%$	$(0,25+1)/(1+1)=62,5\%$	$(0+1)/(1+1)=50\%$

Tabela 18.6: Seleção de alternativa - cálculo de precisão.

Caso	Abrangência		
	ALT1	ALT2	ALT3
1	$(1+1)/(1+1)=100\%$	$(0,4+1)/(2+1)=46,7\%$	$(0+1)/(0+1)=100\%$
2	$(0+1)/(1+1)=50\%$	$(0+1)/(2+1)=33,3\%$	$(0+1)/(0+1)=100\%$
3	$(0,4+1)/(1+1)=70\%$	$(0,35+1)/(2+1)=45\%$	$(0+1)/(0+1)=100\%$
4	$(0,2+1)/(1+1)=60\%$	$(0,5+1)/(2+1)=50\%$	$(0+1)/(0+1)=100\%$
5	$(0,2+1)/(1+1)=60\%$	$(0,5+1)/(2+1)=50\%$	$(0+1)/(0+1)=100\%$
6	$(0,2+1)/(1+1)=60\%$	$(1+1)/(2+1)=66,7\%$	$(0+1)/(0+1)=100\%$
7	$(0,1+1)/(1+1)=55\%$	$(0,25+1)/(2+1)=41,7\%$	$(0+1)/(0+1)=100\%$
8	$(0,3+1)/(1+1)=65\%$	$(0,25+1)/(2+1)=41,7\%$	$(0+1)/(0+1)=100\%$

Tabela 18.7: Seleção de alternativa - cálculo de abrangência.

Caso	Medida F		
	ALT1	ALT2	ALT3
1	$2 \times 1 \times 1 / (1+1) = 1$	$2 \times 0,7 \times 0,467 / (0,7+0,467) = 0,56$	$2 \times 0,5 \times 1 / (0,5+1) = 0,666$
2	$2 \times 1 \times 0,5 / (1+0,5) = 0,66$	$2 \times 1 \times 0,33 / (1+0,33) = 0,5$	$2 \times 1 \times 1 / (1+1) = 1$
3	$2 \times 0,7 \times 0,7 / (0,7+0,7) = 0,7$	$2 \times 0,675 \times 0,45 / (0,675+0,45) = 0,54$	$2 \times 0,5 \times 1 / (0,5+1) = 0,666$
4	$2 \times 0,4 \times 0,6 / (0,4+0,6) = 0,48$	$2 \times 0,33 \times 1 / (1+0,33) = 0,5$	$2 \times 0,5 \times 0,5 / (0,5+0,5) = 0,5$
5	$2 \times 0,4 \times 0,6 / (0,4+0,6) = 0,48$	$2 \times 0,5 \times 0,5 / (0,5+0,5) = 0,5$	$2 \times 0,33 \times 1 / (1+0,33) = 0,5$
6	$2 \times 0,6 \times 0,6 / (0,6+0,6) = 0,6$	$2 \times 1 \times 0,666 / (1+0,666) = 0,8$	$2 \times 0,5 \times 1 / (1+0,5) = 0,667$
7	$2 \times 0,55 \times 0,55 / (0,55+0,55) = 0,55$	$2 \times 0,625 \times 0,417 / (0,625+0,417) = 0,5$	$2 \times 0,5 \times 1 / (1+0,5) = 0,667$
8	$2 \times 0,65 \times 0,65 / (0,65+0,65) = 0,65$	$2 \times 0,625 \times 0,417 / (0,625+0,417) = 0,5$	$2 \times 0,5 \times 1 / (1+0,5) = 0,667$

Tabela 18.8: Seleção de alternativa - cálculo de Medida F.

Caso	Erro Combinado		
	ALT1	ALT2	ALT3
1	0/(0+1)=0%	$(2 \times (1-0,2))/(2+1)=53,3\%$	$1/(1+1)=50\%$
2	$1/(1+1)=50\%$	$(2 \times 1)/(2+1)=66,6\%$	0/(0+1)=0%
3	0,6/(1+1)=30%	$((1-0,1)+(1-0,25))/(2+1)=55,0\%$	$1/(1+1)=50\%$
4	$(2 \times (1-0,1))/(2+1)=60\%$	$(2 \times (1-0,25)+1)/(3+1)=62,5\%$	$2/(2+1)=66,7\%$
5	$(2 \times (1-0,1))/(2+1)=60\%$	$(2 \times (1-0,25))/(2+1)=50\%$	$2/(2+1)=66,7\%$
6	$(1-0,2)/(1+1)=40\%$	$1/(2+1)=33,3\%$	$1/(1+1)=50\%$
7	$(1-0,1)/(1+1)=45\%$	$(1+(1-0,25))/(2+1)=58,3\%$	$1/(1+1)=50\%$
8	$(1-0,3)/(1+1)=35\%$	$(2 \times (1-0,125))/(2+1)=58,3\%$	$1/(1+1)=50\%$

Tabela 18.9: Selecção de alternativa - cálculo de Erro Combinado.

18.3 Tarefa de classificação semântica

A tarefa de classificação semântica avalia até que ponto os sistemas participantes conseguem classificar a EM numa hierarquia de categorias e de tipos definidos no HAREM, que foi especialmente criada o português e foi revista conjuntamente pelos participantes e pela organização.

18.3.1 Medidas

A classificação semântica é avaliada através de quatro medidas, que fornecem mais informação aos participantes sobre o desempenho dos seus sistemas:

Por categorias: pontua-se apenas a categoria da etiqueta.

Por tipos: pontua-se apenas as EM que tiveram categoria(s) pontuada(s) como `correcto`, e onde se avalia somente o atributo `TIPO` da etiqueta.

Combinada: avalia-se as categorias e os tipos da EM, através de uma pontuação que combina as duas através da equação 18.3.

Plana: avalia-se os pares categoria-tipo como folhas de uma classificação plana, considerando apenas como certos os casos que tenham a categoria e o tipo pontuados como `correcto`.

18.3.2 Pontuações

A pontuação na classificação semântica é feita para a categoria e para o tipo, em separado. São usados três valores possíveis:

Correcto: quando a categoria (ou tipo) da EM da saída é igual à categoria (ou tipo) da EM da CD.

Em Falta: quando a categoria (ou tipo) da EM da CD está ausente da categoria (ou tipo) da EM da saída.

Espúrio: quando a categoria (ou tipo) da EM da saída está ausente da categoria (ou tipo) da EM da CD.

Estas são as pontuações usadas para avaliar os alinhamentos, de uma forma genérica. No entanto, como as EM podem ter mais do que uma categoria e tipo (`<ABC... TIPO="XYZ...">`), estas pontuações não podem ser atribuídas assim de uma forma tão linear.

Como tal, vamos detalhar as regras de pontuação para cada medida em separado, ilustrada com exemplos.

Medida por categorias

A pontuação para a classificação semântica medida por categorias avalia as EM da seguinte maneira (ver exemplos na Tabela 18.10):

Caso	Saída Sistema	CD	Correcta	Em Falta	Espúria
1	<A>	<A>	A	-	-
2		<A>	-	A	B
3	<A>	<ABC>	A	-	-
4	<D>	<ABC>	-	A, B e C	D
5	<A>		-	-	A

Tabela 18.10: Pontuação na classificação semântica medida por categorias.

Correcta: Quando o sistema atribui à EM uma categoria, e se essa categoria for igual à da EM na CD, é pontuada como *correcto* (caso 1 da Tabela 18.10). Contudo, se a respectiva EM da CD possui um conjunto de categorias, basta a categoria da EM da saída corresponder a uma desse conjunto, que além de ser pontuado igualmente como *correcto*, o sistema não será prejudicado por faltarem as outras. Ou seja, o caso 3 da Tabela 18.10 resulta na mesma pontuação que o caso 1.

Em Falta: Se a categoria da EM de saída não corresponde à categoria da EM da CD, no caso de esta ter uma classificação única (caso 2 da Tabela 18.10), ou não corresponder a nenhuma das classificações múltiplas (caso 4 da Tabela 18.10), cada uma das categorias da EM da CD é pontuada como *Em Falta*. Contudo, se a categoria que o sistema classificou estiver incluída no conjunto presente na EM da CD, nada é considerado *Em Falta* (caso 3 da Tabela 18.10).

Espúria: no caso da EM do sistema atribuir uma categoria que não existe na EM da CD, essa categoria é pontuada como *espúria* (casos 2, 4 e 5 da Tabela 18.10). Esta marcação é atribuída quer em conjunção com *Em Falta*, quer se o sistema identificou algo como EM que não o seja.

Medida por tipos

Na classificação semântica medida por tipos, as EM são pontuadas de um modo semelhante à da classificação semântica por categorias, mas entrando em conta apenas com os casos em que as categorias foram correctamente identificadas, ou seja, é uma medida relativa por excelência. A Tabela 18.11 resume a pontuação atribuída nos diversos casos. O raciocínio é análogo ao caso anterior referente às categorias.

Caso	Saída Sistema	CD	Correcta	Em Falta	Espúria
1	<A>		-	X	-
2			-	X	-
3		<AAA TIPO="XYZ">	-	X, Y e Z	-
4			X	-	-
5			-	Y	X
6		<ABC TIPO="XYZ">	X	-	-
7		<AAA TIPO="XYZ">	X	-	-
8		<AAA TIPO="WYZ">	-	W, Y e Z	X

Tabela 18.11: Pontuação na classificação semântica medida por tipos.

Medida combinada

A medida semântica combinada combina a pontuação da categoria e do tipo através de uma fórmula única, de modo a indicar o nível da classificação semântica como um todo:

$$P_{CSC} = \begin{cases} 0 & \text{se a categoria não estiver correcta.} \\ 1 & \text{se a categoria estiver correcta mas o tipo não estiver correcto.} \\ 1 + \left(1 - \frac{n_c}{n_t}\right) - \frac{n_e}{n_t} & \text{se a categoria estiver correcta e pelo menos um tipo correcto.} \end{cases} \quad (18.3)$$

Onde n_c representa o número de tipos correctos, n_e o número de tipos espúrios, e n_t o número de tipos possível nessa categoria. Note-se que para calcular estes últimos valores, é preciso naturalmente conhecer quantos tipos diferentes cada categoria pode ter, o que está descrito na Tabela 18.12. Como o número de tipos de certas categorias foram alterados do HAREM para o Mini-HAREM, apresentamos os valores para cada evento:

Categoria	HAREM		Mini-HAREM	
	Número de tipos distintos	Valor máximo	Número de tipos distintos	Valor máximo
ABSTRACCAO	8	1,875	8	1,875
ACONTECIMENTO	3	1,666	3	1,667
COISA	3	1,666	4	1,75
LOCAL	5	1,8	5	1,8
OBRA	4	1,75	3	1,667
ORGANIZACAO	4	1,75	4	1,75
PESSOA	6	1,833	6	1,833
TEMPO	4	1,75	4	1,75
VALOR	3	1,667	3	1,667

Tabela 18.12: Quantidade de tipos distintos que uma categoria semântica pode ter, e valor máximo correspondente para o cálculo da medida combinada, para o HAREM e o Mini-HAREM.

Veja-se a Tabela 18.13 com alguns exemplos, em que assumimos que a categoria *A* tem quatro tipos distintos.

Caso	Saída do Sistema	CD	Medida combinada
1			$1+(1-\frac{0}{4}) = 1$
2			$1+(1-\frac{1}{4}) = 1,75$
3		<AZ TIPO="BY">	$1+(1-\frac{1}{4})-\frac{1}{4} = 1,5$
4		<AA TIPO="CD">	$1+(1-\frac{0}{4}) = 1$

Tabela 18.13: Exemplo para a classificação semântica na medida combinada, para uma categoria *A* com quatro tipos ($n_i = 4$).

Medida plana

Caso	Saída Sistema	CD	Correcta	Em Falta	Espúria
1			(A,X)	-	-
2			-	(A,X)	(A,Y)
3		<AAA TIPO="XYZ">	(A,Y)	-	-
4		<AAA TIPO="XYZ">	(A,X Y Z)	(A,W)	-
5	<B TIPO="Z">		-	(A,X)	(B,Z)

Tabela 18.14: Pontuação da classificação semântica, na medida plana.

A classificação semântica na medida plana tem como objecto de estudo o par (CATEGORIA, TIPO). Por exemplo, se as EM em análise fossem <LOCAL TIPO="GEOGRAFICO">Coimbra</LOCAL> e <PESSOA TIPO="INDIVIDUAL">Magalhães</PESSOA>, então os pares a serem avaliados seriam (LOCAL, GEOGRAFICO) e (PESSOA, INDIVIDUAL), respectivamente. Um par é pontuado como correcto quando a categoria e o tipo são o mesmo na entidade correspondente da CD. A Tabela 18.14 ilustra as regras da medida.

18.3.3 Métricas

Precisão

A precisão apresenta-se sobre dois cenários: absoluto (para todas as EM) e relativo (às EM correctamente identificadas).

Para a medida por categorias, a precisão é dada pela fórmula:

Absoluto: $\text{Precisão}_{\text{medida categorias}} = \frac{(\sum \text{EM correctamente identificadas e com categoria correcta} + Y)}{(\sum \text{EM classificadas pelo sistema})}$

Relativo: $\text{Precisão}_{\text{medida categorias}} = (\sum \text{EM correctamente identificadas e com categoria correcta} + Y) / (\sum \text{EM parcial ou correctamente identificadas classificadas pelo sistema})$

Em que Y corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com categoria correctas. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

A classificação semântica na medida por tipos é, por definição, sempre relativa:

Relativo: $\text{Precisão}_{\text{medida tipos}} = (\sum \text{EM correctamente identificadas e com categoria e tipo correctos} + Z) / (\sum \text{EM correctamente ou parcialmente identificadas})$

Em que Z corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com categoria e tipo correctos. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

Para a classificação semântica combinada, a precisão mede o grau de sucesso de acordo com a classificação máxima (calculada assumindo que todas as categorias e tipos propostos pelo sistema estiverem correctos):

Absoluto: $\text{Precisão}_{\text{medida CSC}} = (\text{Valor de CSC obtida pelo sistema} / \text{Valor máximo da CSC para a saída do sistema})$

Relativo: $\text{Precisão}_{\text{medida CSC}} = (\text{Valor da CSC obtida pelo sistema} / \text{Valor máximo da CSC para a saída do sistema só considerando EM parcial ou correctamente identificadas})$

Para a medida plana, a precisão é calculada da seguinte forma:

Absoluto: $\text{Precisão}_{\text{medida plana}} = (\sum \text{EM correctamente identificadas e com categoria e tipo correctos} + Z) / (\sum \text{EM classificadas pelo sistema})$

Relativo: $\text{Precisão}_{\text{medida plana}} = (\sum \text{EM correctamente identificadas e com categoria e tipo correctos} + Z) / (\sum \text{EM parcial ou correctamente identificadas e classificadas pelo sistema})$

Em que Z corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com categoria e tipo correctos. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

Abrangência

A abrangência define-se de forma diferente para cada uma das quatro medidas, e de forma diferente para os cenários absoluto e relativo.

Para a medida por categorias, a abrangência é calculada da seguinte forma:

Absoluto: $Abrangência_{medida\ categorias} = (\sum EM\ correctamente\ identificadas\ e\ com\ categoria\ correcta + Y) / (\sum EM\ classificadas\ na\ CD)$

Relativo: $Abrangência_{medida\ categorias} = (\sum EM\ correctamente\ identificadas\ e\ com\ categoria\ correcta + Y) / (\sum EM\ parcial\ ou\ correctamente\ identificadas\ e\ classificadas\ na\ CD)$

Em que Y corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com categoria correcta. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

A classificação semântica na medida por tipos é, por definição, sempre relativa:

Relativo: $Abrangência_{medida\ tipos} = (\sum EM\ correctamente\ identificadas\ e\ com\ categoria\ e\ tipo\ correctos + Z) / (\sum EM\ correctamente\ classificadas\ em\ categoria\ na\ CD)$

Em que Z corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com categoria e tipo correctos. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

Na avaliação da classificação semântica combinada, a abrangência mede o nível de cobertura de acordo com a classificação máxima (se tanto as categorias como os tipos enviados estiverem correctos). Mais uma vez, no cenário absoluto usam-se todas as EM na CD, e no relativo apenas o subconjunto parcial ou correctamente identificado.

Absoluto: $Abrangência_{medida\ CSC} = (Valor\ da\ medida\ semântica\ combinada\ obtida\ pelo\ sistema / Valor\ máximo\ da\ medida\ semântica\ combinada\ na\ CD)$

Relativo: $Abrangência_{medida\ CSC} = (Valor\ da\ medida\ semântica\ combinada\ obtida\ pelo\ sistema / Valor\ máximo\ da\ medida\ semântica\ combinada\ na\ CD\ usando\ apenas\ as\ EM\ correctamente\ identificadas)$

Para a medida plana, a abrangência calculada da seguinte forma:

Absoluto: $Abrangência_{medida\ plana} = (\sum EM\ correctamente\ identificadas\ e\ com\ categoria\ e\ tipo\ correctos + Z) / (\sum EM\ na\ CD)$

Relativo: $Abrangência_{medida\ plana} = (\sum EM\ correctamente\ identificadas\ e\ com\ categoria\ e\ tipo\ correctos + Z) / (\sum EM\ parcial\ ou\ correctamente\ identificadas\ na\ CD)$

Em que Z corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com categoria e tipo correctos. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

Sobre-geração

A sobre-geração na classificação semântica mede o número de EM com uma classificação semântica espúria, em comparação com a CD. A sobre-geração é calculada de forma diferente, de acordo com o cenário usado (absoluto ou relativo).

Para a medida por categorias, a sobre-geração é calculada da seguinte forma:

Absoluto: $Sobre-geração_{medida\ categorias} = (\sum EM\ com\ classificação\ semântica\ espúria\ na\ categoria / \sum EM\ classificadas\ com\ categoria\ pelo\ sistema)$

Relativo: $Sobre-geração_{medida\ categorias} = (\sum EM\ parcial\ ou\ correctamente\ identificadas\ com\ classificação\ semântica\ espúria\ na\ categoria) / (\sum EM\ parcial\ ou\ correctamente\ identificadas\ classificadas\ com\ categoria\ pelo\ sistema)$

A classificação semântica na medida por tipos é, por definição, sempre relativa:

Relativo: $Sobre-geração_{medida\ tipos} = (\sum EM\ com\ classificação\ semântica\ espúria\ no\ tipo) / (\sum EM\ parcial\ ou\ correctamente\ identificadas\ classificadas\ com\ categoria\ e\ tipo\ pelo\ sistema)$

A classificação semântica na medida plana é calculada da seguinte forma:

Absoluto: $Sobre-geração_{medida\ plana} = (\sum EM\ com\ classificação\ semântica\ espúria\ na\ categoria\ ou\ no\ tipo) / (\sum EM\ classificadas\ com\ categoria\ e\ tipo\ pelo\ sistema)$

Relativo: $Sobre-geração_{medida\ plana} = (\sum EM\ correctamente\ identificadas\ com\ classificação\ semântica\ espúria\ na\ categoria\ ou\ no\ tipo + W) / (\sum EM\ parcial$

ou correctamente identificadas classificadas com categoria e tipo pelo sistema)

Em que W corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com classificação semântica espúria na categoria ou no tipo. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

Sub-geração

A sub-geração na classificação semântica mede o número de EM com uma classificação semântica em falta, em comparação com a saída. A sub-geração é calculada de forma diferente, de acordo com o cenário usado (absoluto ou relativo).

Para a medida por categorias, a sub-geração é calculada da seguinte forma:

Absoluto: $\text{Sub-geração}_{\text{medida categorias}} = (\sum \text{EM com classificação semântica em falta na categoria}) / (\sum \text{EM com categoria na CD})$

Relativo: $\text{Sub-geração}_{\text{medida categorias}} = (\sum \text{EM correctamente identificadas com classificação semântica em falta na categoria} + R) / (\sum \text{EM parcial ou correctamente identificadas com categoria na CD})$

Em que R corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com classificação semântica em falta na categoria. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

A classificação semântica na medida por tipos é, por definição, sempre relativa:

Relativo: $\text{Sub-geração}_{\text{medida tipos}} = (\sum \text{EM correctamente identificadas com classificação semântica em falta no tipo} + S) / (\sum \text{EM parcial ou correctamente identificadas com tipo na CD})$

Em que S corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com classificação semântica em falta no tipo. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

A classificação semântica na medida plana é calculada da seguinte forma:

Absoluto: $\text{Sub-geração}_{\text{medida plana}} = (\sum \text{EM com classificação semântica em falta na categoria ou no tipo}) / (\sum \text{EM com categoria na CD})$

Relativo: $\text{Sub-geração}_{\text{medida plana}} = (\sum \text{EM correctamente identificadas com classificação semântica em falta na categoria ou no tipo} + T) / (\sum \text{EM parcial ou correctamente identificadas com categoria e tipo na CD})$

Em que T corresponde ao somatório dos valores obtidos para as EM parcialmente identificadas e com classificação semântica em falta na categoria ou no tipo. Esses valores são calculados pela fórmula $\frac{n_c}{n_d}$.

18.3.4 Exemplo detalhado de atribuição de pontuação

Apresentamos um exemplo de texto, etiquetado por um sistema hipotético, e a respectiva CD. Para não sobrecarregar o presente documento, todas as EM da CD são identificadas correctamente ou parcialmente (portanto, os cenários relativo e absoluto produzem os mesmos resultados).

Exemplo da colecção dourada:

Plano hidrológico de <ORGANIZACAO | LOCAL TIPO="ADMINISTRACAO | ADMINISTRATIVO"> Espanha </ORGANIZACAO mid LOCAL> analisado em <LOCAL TIPO="ADMINISTRATIVO"> Lisboa </LOCAL>. Terminou ontem no <LOCAL TIPO="ALARGADO"> Laboratório Nacional de Engenharia Civil </LOCAL>, em <LOCAL TIPO="ADMINISTRATIVO"> Lisboa </LOCAL>, o <ACONTECIMENTO TIPO="EVENTO"> Encontro de Reflexão </ACONTECIMENTO> sobre a concretização do <ABSTRACCAO TIPO="PLANO"> Plano Hidrológico </ABSTRACCAO> espanhol. Em análise esteve um documento que prevê a transferência de significativos volumes de água dos rios <LOCAL TIPO="GEOGRAFICO"> Douro </LOCAL> e <LOCAL TIPO="GEOGRAFICO"> Tejo </LOCAL> para a bacia hidrográfica do rio <LOCAL TIPO="GEOGRAFICO"> Jucar </LOCAL>.

Exemplo de saída do sistema:

<LOCAL TIPO="GEOGRAFICO"> Plano hidrológico de Espanha </LOCAL> analisado em <LOCAL TIPO="ADMINISTRATIVO"> Lisboa </LOCAL>. Terminou ontem no <LOCAL TIPO="ALARGADO"> Laboratório Nacional </LOCAL> de <ORGANIZACAO TIPO="SUB"> Engenharia Civil </ORGANIZACAO>, em <LOCAL TIPO="ADMINISTRATIVO"> Lisboa </LOCAL>, o <ABSTRACCAO TIPO="PLANO"> Encontro de Reflexão </ABSTRACCAO> sobre a

concretização do **<ABSTRACCAO TIPO="PLANO"> Plano Hidrológico**
</ABSTRACCAO> espanhol. **<ABSTRACCAO TIPO="DISCIPLINA">**
Em análise </ABSTRACCAO> esteve um documento que prevê
 a transferência de significativos volumes de água dos rios
<LOCAL TIPO="GEOGRAFICO"> Douro </LOCAL> e **<LOCAL**
TIPO="GEOGRAFICO"> Tejo </LOCAL> para a bacia hidrográfica
 do rio **<ABSTRACCAO TIPO="PLANO"> Jucar </ABSTRACCAO>**.

Nos alinhamentos parciais do exemplo, está associado um factor de correcção (calculado na tarefa de identificação) que influencia os cálculos das várias medidas. Especificamente, referimo-nos aos seguintes casos:

- **<LOCAL TIPO="GEOGRAFICO"> Plano hidrológico de Espanha </LOCAL>**
- **<LOCAL TIPO="ALARGADO"> Laboratório Nacional </LOCAL>**
- **<ORGANIZACAO TIPO="SUB"> Engenharia Civil </ORGANIZACAO>**

Em qualquer destes casos a correspondência com a CD é parcial, logo a sua avaliação tem de ser condicionada por um factor de correcção que condiciona a contribuição desta entidade para a avaliação semântica global. O factor de correcção é dado pela fórmula $\frac{n_c}{n_d}$, onde n_c representa o número de átomos comuns entre as duas EM, e n_d representa o número de átomos distintos entre as duas EM.

Isto significa que a contribuição da EM **<LOCAL TIPO="GEOGRAFICO">Plano hidrológico de Espanha</LOCAL>** é de 0,25 ($n_c = 1$, e $n_d = 4$) e não 1 para o somatório total.

Medida por categorias

Na Tabela 18.15 apresentamos a pontuação para a classificação semântica segundo a medida por categorias, e na Tabela 18.16 os valores das métricas. Note-se que, no caso das identificações parciais, colocamos entre parênteses o correspondente factor de correcção $\frac{n_c}{n_d}$.

Medida por tipos

Na Tabela 18.17 apresentamos a pontuação para a classificação semântica segundo a medida por tipos, e na Tabela 18.18 os valores das métricas. De notar que os casos 4, 6, 8 e 11 da Tabela 18.17 não são classificados, porque não foram pontuados como correctos na Tabela 18.15.

Caso	Saída do Sistema	Correcta	Em Falta	Espúria
1	<LOCAL TIPO="GEOGRAFICO">Plano hidrológico de Espanha</LOCAL>	LOCAL (0,25)	-	-
2	<LOCAL TIPO="ADMINISTRATIVO"> Lisboa</LOCAL>	LOCAL	-	-
3	<LOCAL TIPO="ALARGADO"> Laboratório Nacional</LOCAL>	LOCAL (0,4)	-	-
4	<ORGANIZACAO TIPO="SUB"> Engenharia Civil</ORGANIZACAO>	-	-	ORGANIZACAO
5	<LOCAL TIPO="ADMINISTRATIVO"> Lisboa</LOCAL>	LOCAL	-	-
6	<ABSTRACCAO TIPO="PLANO"> Encontro de Reflexão</ABSTRACCAO>	-	ACONTECIMENTO	ABSTRACCAO
7	<ABSTRACCAO TIPO="PLANO">Plano Hidrológico</ABSTRACCAO>	ABSTRACCAO	-	-
8	<ABSTRACCAO TIPO="DISCIPLINA"> Em análise</ABSTRACCAO>	-	-	ABSTRACCAO
9	<LOCAL TIPO="GEOGRAFICO"> Douro</LOCAL>	LOCAL	-	-
10	<LOCAL TIPO="GEOGRAFICO"> Tejo</LOCAL>	LOCAL	-	-
11	<ABSTRACCAO TIPO="PLANO"> Jucar</ABSTRACCAO>	-	LOCAL	ABSTRACCAO
Total		5,65	2	4

Nota: No caso 4, como a EM anterior do sistema alinhou com a mesma EM da CD, e foi pontuada como correcta no alinhamento anterior, não podemos pontuar a categoria LOCAL como em falta.

Tabela 18.15: Pontuação da classificação semântica medida por categorias, para o exemplo dado.

Métrica	Valor
Precisão	$\frac{5,65}{11} = 51,36\%$
Abrangência	$\frac{5,65}{9} = 62,77\%$
Medida F	$\frac{2 \times 0,5136 \times 0,6277}{0,5136 + 0,6277} = 0,565$
Sobre-geração	$\frac{4}{11} = 36,36\%$
Sub-geração	$\frac{2}{9} = 22,2\%$

Tabela 18.16: Valores das métricas para a tarefa de classificação semântica, medida por categorias, para o exemplo dado.

Medida combinada

Na Tabela 18.19 apresentamos a pontuação para a classificação semântica segundo a medida combinada, e na Tabela 18.20 os valores das métricas. Salientamos que os casos 1 e 3

Caso	Saída do Sistema	Correcta	Em Falta	Espúria
1	<LOCAL TIPO="GEOGRAFICO"> Plano hidrológico de Espanha</LOCAL>	-	ADMINISTRATIVO	GEOGRAFICO
2	<LOCAL TIPO="ADMINISTRATIVO"> Lisboa</LOCAL>	ADMINISTRATIVO	-	-
3	<LOCAL TIPO="ALARGADO"> Laboratório Nacional</LOCAL>	ALARGADO (0,4)	-	-
5	<LOCAL TIPO="ADMINISTRATIVO"> Lisboa</LOCAL>	ADMINISTRATIVO	-	-
7	<ABSTRACCAO TIPO="PLANO"> Plano Hidrológico</ABSTRACCAO>	PLANO	-	-
9	<LOCAL TIPO="GEOGRAFICO"> Douro</LOCAL>	GEOGRAFICO	-	-
10	<LOCAL TIPO="GEOGRAFICO"> Tejo</LOCAL>	GEOGRAFICO	-	-
Total		5,4	1	1

Tabela 18.17: Pontuação da classificação semântica por tipos, para o exemplo dado.

Métrica	Valor
Precisão	$\frac{5,4}{7} = 77,14\%$
Abrangência	$\frac{5,4}{7} = 77,14\%$
Medida F	$\frac{2 \times 0,7714 \times 0,7714}{0,7714 + 0,7714} = 0,7714$
Sobre-geração	$\frac{1}{7} = 14,28\%$
Sub-geração	$\frac{1}{7} = 14,28\%$

Tabela 18.18: Valores das métricas para a classificação semântica, medida por tipos, para o exemplo dado.

da Tabela 18.19 são multiplicados pelo factor de correcção $\frac{n_c}{n_d}$, respectivamente, 0,25 e 0,4.

Medida plana

Na Tabela 18.21 apresentamos a pontuação para a classificação semântica segundo a medida plana, e na Tabela 18.22 os valores das métricas. Salientamos que os casos 1 e 3 da Tabela 18.21 são multiplicados pelo factor de correcção, 0,25 e 0,4, respectivamente.

Caso	Classificação
1	$1+0 \times \left(1 - \frac{1}{5}\right) \times 0,25 = 0,25$
2	$1+1 \times \left(1 - \frac{1}{5}\right) = 1,80$
3	$1+1 \times \left(1 - \frac{1}{5}\right) \times 0,4 = 0,72$
4	0,0
5	$1+1 \times \left(1 - \frac{1}{5}\right) = 1,80$
6	0,0
7	$1+1 \times \left(1 - \frac{1}{8}\right) = 1,875$
8	0,0
9	$1+1 \times \left(1 - \frac{1}{5}\right) = 1,80$
10	$1+1 \times \left(1 - \frac{1}{5}\right) = 1,80$
11	0,0
Total	10,045

Tabela 18.19: Pontuação da classificação semântica segundo a medida combinada, para o exemplo dado.

Métrica	Valor
Precisão máxima do sistema	$\frac{10,045}{20,05} = 50,1\%$
Abrangência Máxima na CD	$\frac{10,045}{16,14} = 62,2\%$
Medida F	$\frac{2 \times 0,501 \times 0,6223}{0,501 + 0,6223} = 0,555$

Nota: o denominador do cálculo da precisão máxima do sistema corresponde ao somatório do cálculo da classificação semântica combinada assumindo que as classificações atribuídas pelo sistema estão totalmente correctas. Para melhor perceber este conceito imagine que as categorias da Tabela 18.15 e os tipos (agora com os restantes casos 4, 6, 8 e 11) da Tabela 18.17 estivessem a ser sempre considerados correctos. Analogamente, o denominador do cálculo da abrangência máxima da CD utiliza a mesma fórmula para calcular o somatório das classificações combinadas para cada uma das entidades na CD.

Tabela 18.20: Valores das métricas para a tarefa de classificação semântica, segundo a medida combinada, para o exemplo dado.

Caso	Saída do Sistema	Correcta	Em Falta	Espúria
1	<LOCAL TIPO="GEOGRAFICO">Plano hidrológico de Espanha</LOCAL>	-	(LOCAL, ADMINISTRATIVO)	(LOCAL, GEOGRAFICO)
2	<LOCAL TIPO="ADMINISTRATIVO">Lisboa</LOCAL>	(LOCAL, ADMINISTRATIVO)	-	-
3	<LOCAL TIPO="ALARGADO">Laboratório Nacional</LOCAL>	(LOCAL, ALARGADO)(0.4)	-	-
4	<ORGANIZACAO TIPO="SUB">Engenharia Civil</ORGANIZACAO>	-	* (ORGANIZACAO, SUB)	-
5	<LOCAL TIPO="ADMINISTRATIVO">Lisboa</LOCAL>	(LOCAL, ADMINISTRATIVO)	-	-
6	<ABSTRACCAO TIPO="PLANO">Encontro de Reflexão</ABSTRACCAO>	-	(ACONTECIMENTO, EVENTO)	(ABSTRACCAO, PLANO)
7	<ABSTRACCAO TIPO="PLANO">Plano Hidrológico</ABSTRACCAO>	(ABSTRACCAO, PLANO)	-	-
8	<ABSTRACCAO TIPO="DISCIPLINA">Em análise</ABSTRACCAO>	-	-	(ABSTRACCAO, DISCIPLINA)
9	<LOCAL TIPO="GEOGRAFICO">Douro</LOCAL>	(LOCAL, GEOGRAFICO)	-	-
10	<LOCAL TIPO="GEOGRAFICO">Tejo</LOCAL>	(LOCAL, GEOGRAFICO)	-	-
11	<ABSTRACCAO TIPO="PLANO">Jucar</ABSTRACCAO>	-	(LOCAL, GEOGRAFICO)	(ABSTRACCAO, PLANO)
Total	-	5,4	3	5

Tabela 18.21: Valores das métricas para a classificação semântica, segundo a medida plana, para o exemplo dado.

Métrica	Valor
Precisão	$\frac{5,4}{11} = 49,09\%$
Abrangência	$\frac{5,4}{9} = 60,00\%$
Medida F	$\frac{2 \times 0,4909 \times 0,6000}{0,4909 + 0,6000} = 0,5400$
Sobre-geração	$\frac{5}{11} = 45,45\%$
Sob-geração	$\frac{3}{9} = 33,33\%$

Tabela 18.22: Avaliação global da tarefa de classificação semântica segundo a medida plana.

18.4 Tarefa de classificação morfológica

A tarefa de classificação morfológica tem por objectivo avaliar a aptidão do sistema em definir qual o género e o número das EM identificadas, em comparação com as respectivas classificações morfológicas feitas manualmente na CD.

18.4.1 Medidas

A tarefa de classificação morfológica é avaliada segundo três medidas:

número: só é considerada a pontuação relativamente ao número.

género: só é considerada a pontuação relativamente ao género.

combinada: combina-se as pontuações para género e para o número.

Note-se, além disso, que a avaliação morfológica é apenas feita sobre as EM que também foram classificadas morfológicamente na CD. As classificações morfológicas feitas a EM que não estão classificadas na CD (como por exemplo as EM de categoria *TEMPO*) são simplesmente ignoradas no processamento subsequente.

18.4.2 Pontuações

As pontuações na tarefa de classificação morfológica podem variar de acordo com o cenário de avaliação usado. Em certos casos como é ilustrado no caso 10 da Tabela 18.23, podemos constatar que a pontuação no cenário absoluto é espúria, enquanto que no cenário relativo, a EM é ignorada para efeitos de pontuação. Tal facto deve-se ao facto de, no cenário relativo, as EM que são espúrias na tarefa de identificação também são ignoradas na tarefa de classificação morfológica.

Quando uma EM é imperfeitamente reconhecida (ou seja, foi classificada na tarefa de identificação como parcialmente correcta), apenas contamos os casos em que essa identificação parcial concordava na primeira palavra da EM, multiplicando por um peso de 0,5 as EM que estão morfológicamente correctas.

A pontuação para cada uma das medidas segue as regras ilustradas na Tabela 18.23.

Nas tabelas seguintes, vamos mais uma vez considerar, para simplicidade de exposição, que os exemplos são relativos a EM que o participante queria classificar (cenário selectivo), ou então a todas as etiquetas da CD (cenário total), e que todas as identificações estavam correctas.

Se estivermos num cenário relativo (ou seja, só considerando as EM com valor de pontuação maior que 0 na tarefa de identificação) e os 10 exemplos da Tabela 18.23 como um exemplo de saída do sistema participante (note-se que os casos 9 e 10 serão ignorados e não contabilizados), a avaliação global produziria os resultados apresentados na Tabela 18.24.

Caso	Classificação		Medida		
	CD	Sistema	Gênero	Número	Combinada
1	M,S	M,S	Correcto	Correcto	Correcto
2	M,S	F,S	Incorrecto	Correcto	Incorrecto
3	M,S	M,P	Correcto	Incorrecto	Incorrecto
4	M,S	F,P	Incorrecto	Incorrecto	Incorrecto
5	M,S	?,S	Em Falta	Correcto	Em Falta
6	?,S	M,S	Sobre-especificado	Correcto	Incorrecto
7	?,S	?,S	Correcto	Correcto	Correcto
8	M,S	Não submetido	Em Falta	Em Falta	Em Falta
9	sem identificação	Não submetido	Ignorado	Ignorado	Ignorado
10	sem identificação	Submetido, sem ter ?	(Cen. Relativo)	(Cen. Relativo)	(Cen. Relativo)
			(Cen. Absoluto)	(Cen. Absoluto)	(Cen. Absoluto)

Tabela 18.23: Pontuação para a classificação morfológica, segundo as três medidas.

Métrica	Cenário Absoluto		
	Gênero	Número	Combinada
Precisão	$\frac{3}{8} = 37,5\%$	$\frac{5}{8} = 62,5\%$	$\frac{2}{8} = 25,0\%$
Abrangência	$\frac{3}{8} = 37,5\%$	$\frac{5}{8} = 62,5\%$	$\frac{2}{8} = 25,0\%$
Medida F	$\frac{2 \times 0,375 \times 0,375}{(0,375 + 0,375)} = 0,375$	$\frac{2 \times 0,625 \times 0,625}{(0,625 + 0,625)} = 0,625$	$\frac{2 \times 0,25 \times 0,25}{(0,25 + 0,25)} = 0,25$
Sobre-especificação	$\frac{1}{8} = 12,5\%$	$\frac{0}{8} = 0\%$	-
Sub-geração	$\frac{2}{8} = 25,0\%$	$\frac{1}{8} = 12,5\%$	-

Métrica	Cenário Relativo		
	Gênero	Número	Combinada
Precisão	$\frac{3}{7} = 42,8\%$	$\frac{5}{7} = 71,4\%$	$\frac{2}{7} = 28,3\%$
Abrangência	$\frac{3}{8} = 37,5\%$	$\frac{5}{8} = 62,5\%$	$\frac{2}{8} = 25,0\%$
Medida F	$\frac{2 \times 0,428 \times 0,375}{(0,428 + 0,375)} = 0,40$	$\frac{2 \times 0,714 \times 0,625}{(0,714 + 0,625)} = 0,666$	$\frac{2 \times 0,283 \times 0,25}{(0,283 + 0,25)} = 0,266$
Sobre-especificação	$\frac{1}{7} = 14,3\%$	$\frac{0}{7} = 0\%$	-
Sub-geração	$\frac{2}{8} = 25,0\%$	$\frac{1}{8} = 12,5\%$	-

Tabela 18.24: Valor das métricas para as três medidas da classificação morfológica, considerando os 10 casos da Tabela 18.23.

18.4.3 Métricas

Precisão

Na tarefa de classificação morfológica, a precisão mede o teor de classificações em género/número correctas de todas as produzidas pelo sistema (que tenham classificação morfológica na CD). Ou seja, excluindo sempre os casos em que a EM da CD não se encontra marcada morfológicamente.

Apresentamos a precisão para as três medidas (género, número e combinada), e para os dois cenários de avaliação: independente da identificação (absoluto), ou apenas para os casos em que a identificação obteve pontuação correcta ou parcialmente correcta (relativo).

Absoluto: $\text{Precisão}_{\text{género}} = (\sum \text{EM identificadas correctamente e com género correcto} + 0,5\sum \text{EM identificadas parcialmente correctamente e com género correcto}) / (\sum \text{EM com classificações de género produzidas pelo sistema})$

Relativo: $\text{Precisão}_{\text{género}} = (\sum \text{EM identificadas correctamente e com género correcto} + 0,5\sum \text{EM identificadas parcialmente correctamente e com género correcto}) / (\sum \text{EM com classificações de género produzidas pelo sistema em EM identificadas correctamente ou parcialmente})$

Absoluto: $\text{Precisão}_{\text{número}} = (\sum \text{EM identificadas correctamente e com número correcto} + 0,5\sum \text{EM identificadas parcialmente correctamente e com número correcto}) / (\sum \text{EM com classificações de número produzidas pelo sistema})$

Relativo: $\text{Precisão}_{\text{número}} = (\sum \text{EM identificadas correctamente e com número correcto} + 0,5\sum \text{EM identificadas parcialmente correctamente e com número correcto}) / (\sum \text{EM com classificações de número produzidas pelo sistema em EM identificadas correctamente ou parcialmente})$

Absoluto: $\text{Precisão}_{\text{combinada}} = (\sum \text{EM identificadas correctamente e com género e número correcto} + 0,5\sum \text{EM identificadas parcialmente correctamente e com género e número correcto}) / (\sum \text{EM com classificações de número e género produzidas pelo sistema})$

Relativo: $\text{Precisão}_{\text{combinada}} = (\sum \text{EM identificadas correctamente e com género e número correcto} + 0,5\sum \text{EM identificadas parcialmente correctamente e com género e número correcto}) / (\sum \text{EM com classificações de número e género produzidas pelo sistema em EM identificadas correctamente ou parcialmente})$

Abrangência

Na tarefa de classificação morfológica, a abrangência mede o teor de classificações em género/número que se encontram na CD e que o sistema conseguiu acertar. Tal como para a precisão, mede-se a abrangência no género morfológico, no número morfológico, e na combinação de ambos. No cenário relativo, restringe-se o denominador às EM da CD que foram parcial ou correctamente identificadas pelo sistema.

Absoluto: $Abrangência_{género} = (\sum \text{EM correctamente identificadas com classificações de género correctas} + 0,5\sum \text{EM identificadas parcialmente correctamente com classificações de género correctas}) / (\sum \text{EM com classificações de género na CD})$

Relativo: $Abrangência_{género} = (\sum \text{EM correctamente identificadas com classificações de género correctas} + 0,5\sum \text{EM identificadas parcialmente correctamente com classificações de género correctas}) / (\sum \text{EM parcial ou correctamente identificadas com classificações de género na CD})$

Absoluto: $Abrangência_{número} = (\sum \text{EM correctamente identificadas com classificações de número correctas} + 0,5\sum \text{EM identificadas parcialmente correctamente com classificações de número correctas}) / (\sum \text{EM com classificações de número na CD})$

Relativo: $Abrangência_{número} = (\sum \text{EM correctamente identificadas com classificações de número correctas} + 0,5\sum \text{EM identificadas parcialmente correctamente com classificações de número correctas}) / (\sum \text{EM parcial ou correctamente identificadas com classificações de número na CD})$

Absoluto: $Abrangência_{combinada} = (\sum \text{EM correctamente identificadas com classificações de número e género correctas} + 0,5\sum \text{EM identificadas parcialmente correctamente com classificações de número e género correctas}) / (\sum \text{EM com classificação morfológica na CD})$

Relativo: $Abrangência_{combinada} = (\sum \text{EM correctamente identificadas com classificações de número e género correctas} + 0,5\sum \text{EM identificadas parcialmente correctamente com classificações de número e género correctas}) / (\sum \text{EM parcial ou correctamente identificadas com classificação morfológica na CD})$

Note-se que os denominadores para as três medidas (género, número e combinada), embora formulados de maneira diferente, são exactamente iguais.

Sobre-geração

Relembramos que não se considera, para efeitos de avaliação, espúrios morfológicos (ou seja, só contam para avaliação os casos que também contêm classificação morfológica na CD). Assim, só no cenário absoluto é que há medida de sobre-geração, uma vez que num cenário relativo, não existem EM com morfologia identificadas como espúrias, sendo portanto o valor desta medida sempre 0.

Absoluto: $\text{Sobre-geração}_{\text{género}} = (\sum \text{EM com classificações em género espúrias}) / (\sum \text{EM com classificações em género produzidas pelo sistema e que tenham também classificação morfológica na CD})$

Absoluto: $\text{Sobre-geração}_{\text{número}} = (\sum \text{EM com classificações em número espúrias}) / (\sum \text{EM com classificações de número produzidas pelo sistema e que tenham também classificação morfológica na CD})$

Absoluto: $\text{Sobre-geração}_{\text{combinada}} = (\sum \text{EM com classificações em número ou género espúrias}) / (\sum \text{EM com classificações de número ou género produzidas pelo sistema e que tenham também classificação morfológica na CD})$

Sobre-especificação

Para a tarefa de classificação morfológica, consideramos também a medida de sobre-especificação, que mede a percentagem dos casos sobre-especificados em todos os casos analisados pelo sistema. Por sobre-especificado entendemos os casos em que na CD está "?" e o sistema escolheu um determinado valor concreto.

Absoluto: $\text{Sobre-especificação}_{\text{género}} = (\sum \text{EM com classificações de género sobre-especificadas em EM identificadas correctamente} + 0,5\sum \text{EM com classificações em género sobre-especificadas em EM identificadas parcialmente correctamente}) / (\sum \text{EM com classificações de género produzidas pelo sistema})$

Relativo: $\text{Sobre-especificação}_{\text{género}} = (\sum \text{EM com classificações de género sobre-especificadas em EM identificadas correctamente} + 0,5\sum \text{EM com classificações em género sobre-especificadas em EM identificadas parcialmente correctamente}) / (\sum \text{EM com classificações de género produzidas pelo sistema em EM identificadas parcial ou correctamente})$

Absoluto: Sobre-especificação_{número} = $(\sum \text{EM com classificações de número sobre-especificadas em EM identificadas correctamente} + 0,5\sum \text{EM com classificações em número sobre-especificadas em EM identificadas parcialmente correctamente}) / (\sum \text{EM com classificações de número produzidas pelo sistema})$

Relativo: Sobre-especificação_{número} = $(\sum \text{EM com classificações de número sobre-especificadas em EM identificadas correctamente} + 0,5\sum \text{EM com classificações em número sobre-especificadas em EM identificadas parcialmente correctamente}) / (\sum \text{EM com classificações de número produzidas pelo sistema em EM identificadas parcial ou correctamente})$

Absoluto: Sobre-especificação_{combinada} = $(\sum \text{EM com classificações de número ou género sobre-especificadas em EM identificadas correctamente} + 0,5\sum \text{EM com classificações em número ou género sobre-especificadas em EM identificadas parcialmente correctamente}) / (\sum \text{EM com classificações morfológicas produzidas pelo sistema})$

Relativo: Sobre-especificação_{combinada} = $(\sum \text{EM com classificações de número ou género sobre-especificadas em EM identificadas correctamente} + 0,5\sum \text{EM com classificações em número ou género sobre-especificadas em EM identificadas parcialmente correctamente}) / (\sum \text{EM com classificações morfológicas produzidas pelo sistema em EM identificadas parcial ou correctamente})$

Sub-geração

Na tarefa de classificação morfológica, a subgeração mede o número de classificações em falta comparadas com a informação morfológica na CD. Classificações em falta incluem tanto casos em que nenhuma classificação foi dada, como casos em que o sistema pôs ? para a classificação do género ou número enquanto na CD existe um valor mais específico. Como anteriormente, apresentamos separadamente as fórmulas para o cenário absoluto e relativo.

Absoluto: Sub-geração_{género} = $(\sum \text{EM com classificações em género em falta}) / (\sum \text{classificações em género na CD})$

Relativo: Sub-geração_{género} = $(\sum \text{EM parcial ou correctamente identificadas com classificações em género em falta}) / (\sum \text{EM parcial ou correctamente identificadas com classificações em género na CD})$

Absoluto: $\text{Sub-geração}_{\text{número}} = (\sum \text{EM com classificações em número em falta}) / (\sum \text{classificações em número na CD})$

Relativo: $\text{Sub-geração}_{\text{número}} = (\sum \text{EM parcial ou correctamente identificadas com classificações em número em falta}) / (\sum \text{EM parcial ou correctamente identificadas com classificações em número na CD})$

Absoluto: $\text{Sub-geração}_{\text{combinada}} = (\sum \text{EM com classificações em género ou número em falta}) / (\sum \text{classificações morfológicas na CD})$

Relativo: $\text{Sub-geração}_{\text{combinada}} = (\sum \text{EM parcial ou correctamente identificadas com classificações em género em falta}) / (\sum \text{EM parcial ou correctamente identificadas com classificações morfológicas na CD})$

18.5 Apresentação dos resultados

Os resultados da avaliação são depois apresentados sob duas formas:

Globais: centrados sobre os diversos aspectos da avaliação (por uma determinada categoria, um cenário ou um género textual, por exemplo). Aqui, o desempenho das várias saídas (devidamente anonimizadas) são reunidas em torno de tabelas e/ou gráficos, para permitir uma análise global sobre o comportamento dos sistemas para cada aspecto da avaliação.

Individuais: centrado sobre o desempenho de uma saída. As tabelas e/ou gráficos mostram a posição que a saída ocupou em relação às restantes saídas (devidamente anonimizadas). Estes relatórios possuem dados adicionais sobre o desempenho da saída que não são usados nos relatórios globais.

18.5.1 Resultados globais

Para os resultados globais, apresentam-se várias tabelas comparativas do desempenho dos sistemas. Cada tabela diz respeito a um conjunto dos seguintes parâmetros:

Tarefa: pode ser identificação, classificação morfológica ou classificação semântica.

Por critérios: pode ser global, ou discriminado por categorias, por género textual ou por variante.

Cenário: pode ser total (absoluto ou relativo, nas tarefas de classificação) ou selectivo (absoluto ou relativo, nas tarefas de classificação).

Medida: gênero. número ou combinada (classificação morfológica), ou por categorias, por tipos, combinada ou plana (na classificação semântica).

De reparar que, nos relatórios globais, os sistemas são devidamente anonimizados, tendo os nomes das saídas sido substituídos por pseudónimos.

As tabelas apresentam os valores para as métricas para cada medida / cenário usado. Um exemplo de tabela, para a tarefa de identificação, no global, em cenário total, é assim representada:

	Precisão (%)	Abrangência (%)	Medida F	Erro combinado	Sobre-geração	Sub-geração
riad	78,50	82,84	0,8061	0,2752	0,07913	0,07329
casablanca	77,15	84,35	0,8059	0,2721	0,09134	0,03575
ancara	76,85	83,56	0,8006	0,2781	0,08966	0,04035
sana	77,43	69,57	0,7329	0,3796	0,09524	0,2079
bahreïn	59,45	64,39	0,6182	0,5056	0,2018	0,1607
asmara	56,95%	64,39%	0,6044	0,5230	0,2353	0,1607

Tabela 18.25: Exemplo de uma tabela no relatório global, que compara o desempenho de várias saídas para uma determinada tarefa.

Nos relatórios globais, a tabela é acompanhada também de gráficos. Os valores são apresentados em forma de gráfico de barras (ver Figura 18.1) e em forma de gráfico de pontos (ver Figura 18.2). Nos gráficos de barras, as saídas ficam no eixo das ordenadas, e nos gráficos de pontos, cada ponto representa uma saída.

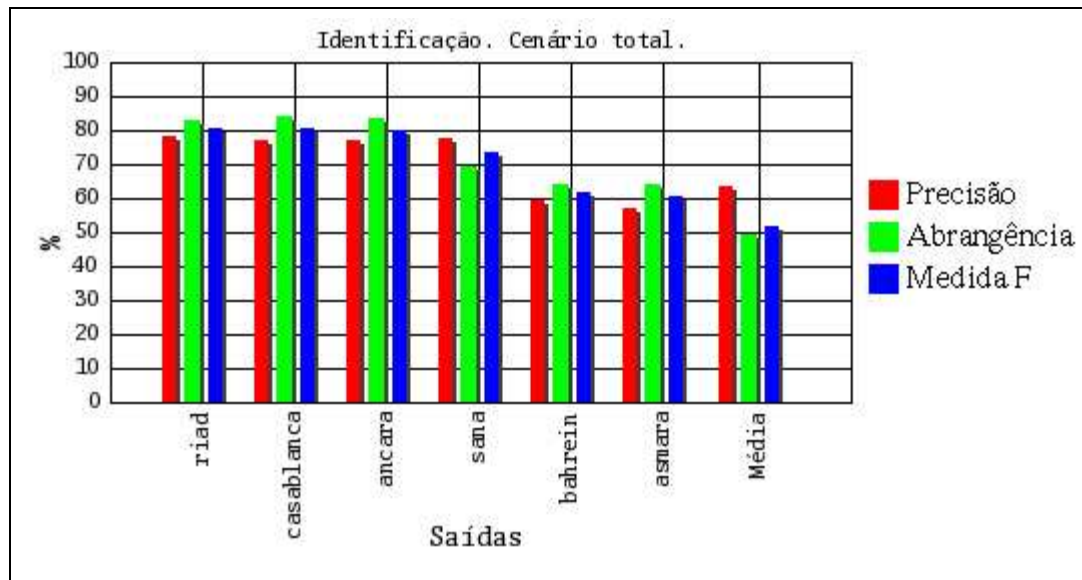


Figura 18.1: Exemplo de um gráfico de barras para o relatório global da tarefa de identificação (cenário total), apresentando os valores da precisão, abrangência e medida F.

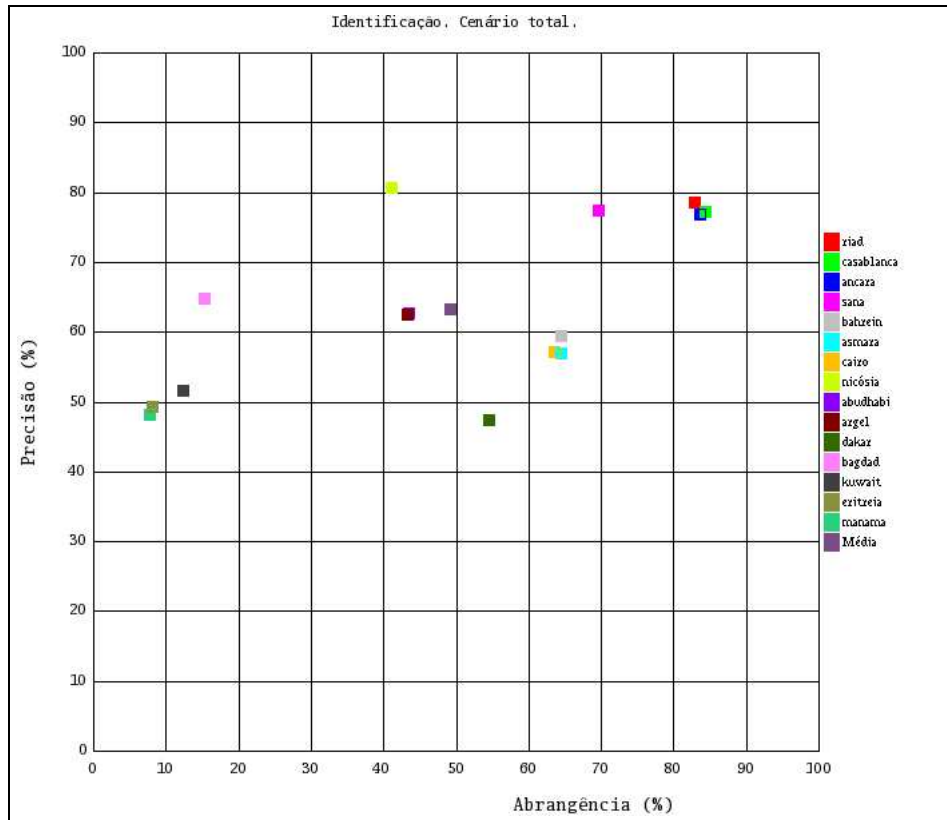


Figura 18.2: Exemplo de um gráfico de pontos para o relatório global da tarefa de identificação (cenário total).

18.5.2 Resultados individuais

Os resultados individuais de cada saída são gerados pelo módulo Alcaide (descrito em 19.2.15, na secção 19.2.15) com base nos relatórios globais, mas com os seguintes melhoramentos:

Resultados filtrados: Nas tabelas de resultados, só se mostra o desempenho das saídas do sistema. A tabela é complementada com informação adicional dos valores de avaliação detalhados. Nos respectivos gráficos de barras, mostra-se também o desempenho de todas as saídas, mas na legenda mostra-se o nome real das saídas do sistema, em vez dos respectivos pseudónimos. Note-se um exemplo de desempenho, para a saída do sistema RENA, na Tabela 18.26 e Figuras 18.3 e 18.4.

Agrupamento de cenários: Enquanto que nos relatórios globais, os resultados são discriminados por cada item (ou seja, há uma tabela para os desempenhos para cada ca-

Total na CD: 5002. Identificadas: 4494. Correctas: 3305 (66,07%).

Parcialmente Correctas: 836 (16,71%). Espúrias: 428 (8,56%). Em Falta: 1040 (20,79%).

Posição	Precisão (%)	Abrangência (%)	Medida F	Erro combinado	Sobre-geração	Sub-geração
4º	77,43	69,57	0,7329	0,3796	0,09524	0,2079

Tabela 18.26: Tabela do relatório individual para a saída RENA, para a tarefa de identificação.

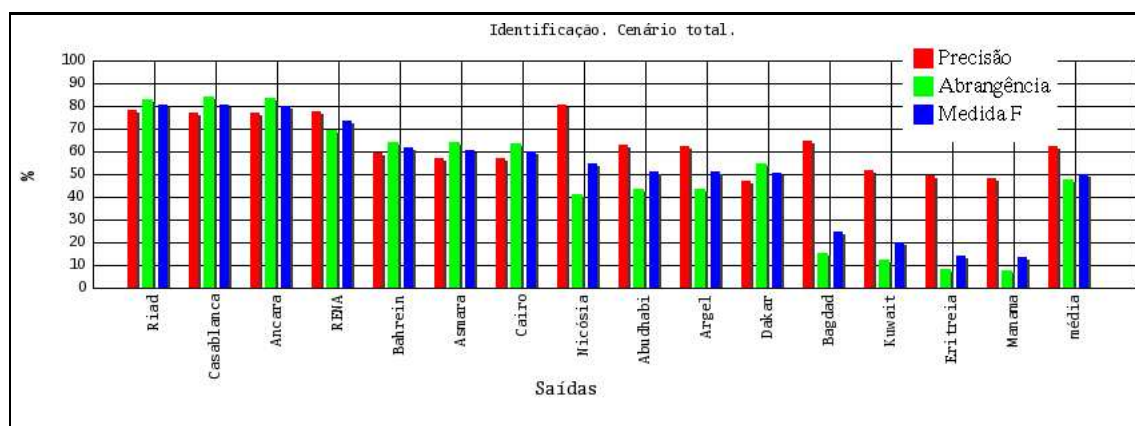


Figura 18.3: Exemplo de um gráfico de barras para o relatório individual da tarefa de identificação (cenário total) para a saída RENA, apresentando os valores da precisão, abrangência e medida F.

tegoria, género textual ou variante), nos relatórios individuais os desempenhos da saída são reunidos numa única tabela. O nome da saída é substituído pela sua posição relativa às outras saídas. Adicionalmente, os valores de avaliação detalhados são agrupados também em tabelas novas (ver Tabelas 18.27 e 18.28).

Gráficos de pontos individual: No caso de cenários (categoria, género textual ou variante), os gráficos de pontos apresentam o desempenho da saída para cada item, em vez de comparar para as restantes saídas como no relatório global (ver Figura 18.5).

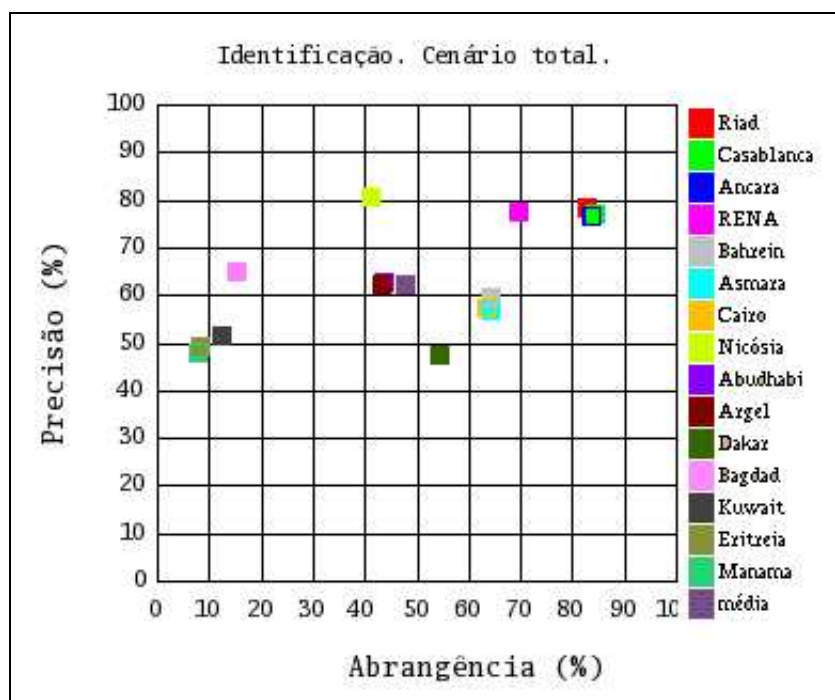


Figura 18.4: Exemplo de um gráfico de pontos para o relatório individual da tarefa de identificação (cenário total) para a saída RENA.

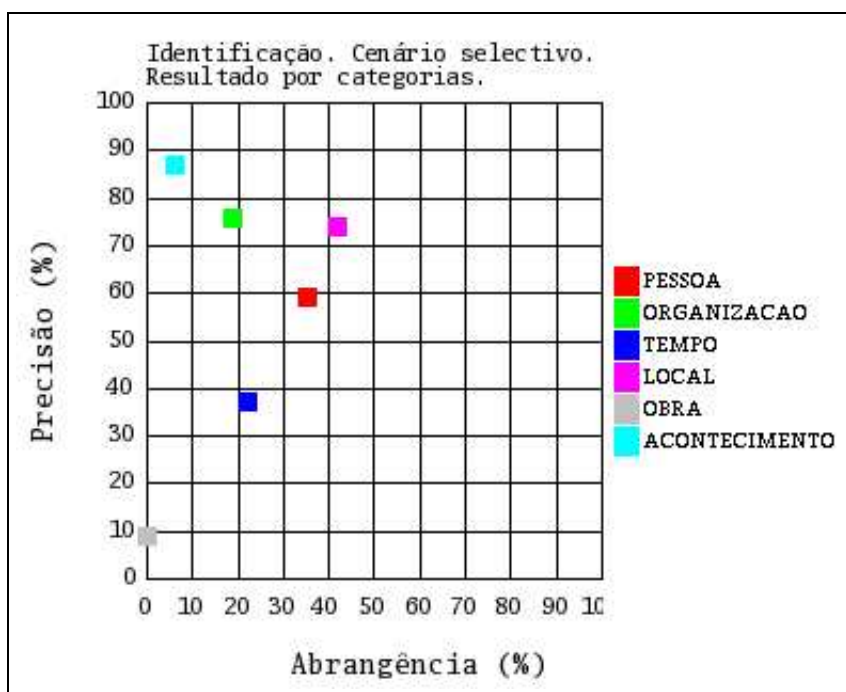


Figura 18.5: Exemplo de um gráfico de pontos para o relatório individual da tarefa de identificação (cenário total) para a saída RENA, discriminada por categorias.

Categoria	Total CD	Identificadas	Correctas		Parc. correctas		Espúrias		Em Falta	
			Total	%	Total	%	Total	%	Total	%
PESSOA	1024	619	339	33,11%	108	10,55%	178	17,38%	580	56,64%
ORGANIZACAO	955	242	176	18,43%	33	3,46%	36	3,77%	746	78,12%
TEMPO	434	264	96	22,12%	11	2,53%	161	37,10%	327	75,35%
LOCAL	1244	713	521	41,88%	47	3,78%	145	11,66%	678	54,50%
OBRA	215	4	0	0,00%	1	0,47%	3	1,40%	214	99,53%
ACONTECIMENTO	109	8	7	6,42%	0	0,00%	1	0,92%	102	93,58%
ABSTRACCAO	453	0	0	0,00%	0	0,00%	0	0,00%	453	100,00%
COISA	81	0	0	0,00%	0	0,00%	0	0,00%	81	100,00%
VALOR	479	0	0	0,00%	0	0,00%	0	0,00%	479	100,00%

Tabela 18.27: Exemplo de uma tabela com valores de avaliação detalhados do relatório individual. No caso presente, os valores referem-se aos desempenhos da saída RENA para a tarefa de identificação, discriminadas por categorias (cenário total).

Categoria	Posição	Precisão (%)	Abrangência (%)	Medida F	Erro combinado	Sobre-geração	Sub-geração
PESSOA	5º	59,23	35,80	0,4463	0,6958	0,2876	0,5664
ORGANIZACAO	7º	76,03	19,27	0,3074	0,8143	0,1488	0,7812
TEMPO	7º	37,44	22,77	0,2832	0,8339	0,6098	0,7535
LOCAL	7º	74,55	42,73	0,5432	0,6179	0,2034	0,5450
OBRA	5º	9,375	0,1744	0,003425	0,9983	0,7500	0,9953
ACONTECIMENTO	5º	87,50	6,422	0,1197	0,9364	0,1250	0,9358

Tabela 18.28: Exemplo de uma tabela de desempenho discriminado do relatório individual. No caso presente, os valores referem-se aos desempenhos da saída RENA para a tarefa de identificação, discriminadas por categorias (cenário total).