

Capítulo 17

Directivas para a identificação e classificação morfológica na colecção dourada do HAREM

Nuno Cardoso, Diana Santos e Rui Vilela

Este capítulo foi previamente publicado como Relatório Técnico DI/FCUL TR-06-19, 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 17, p. 239–244, 2007.

Neste documento, apresentamos as directivas usadas na tarefa de classificação morfológica da colecção dourada do HAREM e, conseqüentemente, qual o comportamento esperado pelos sistemas que participem na tarefa. No capítulo 16 foi indicada a metodologia seguida na classificação semântica.

O texto deste capítulo é exactamente idêntico ao do capítulo anterior, por isso foi aqui omitido. A razão dessa replicação deve-se ao facto de ter sido possível, no HAREM, de participar exclusivamente numa das duas tarefas de classificação. Como tal, e para que os participantes numa e não noutra tarefa não tivessem que ler as duas directivas, optou-se por repetir a secção respectiva às directivas de identificação, que neste capítulo, foi omitida.

17.1 Regras gerais da tarefa de classificação morfológica

Considerámos como passíveis de ser classificadas morfológicamente (isto é, EM que devem ter o atributo MORF):

- As categorias PESSOA, ORGANIZACAO, COISA, ABSTRACCAO, ACONTECIMENTO, OBRA, e VARIADO na sua totalidade.
- Na categoria LOCAL, os tipos ADMINISTRATIVO e GEOGRAFICO.
- Na categoria TEMPO, o tipo CICLICO.

As seguintes EM não têm atributo MORF:

- A categoria VALOR na sua totalidade.
- Na categoria LOCAL, os tipos CORREIO.
- Na categoria TEMPO, o tipo HORA.

E finalmente, nos seguintes casos as EM podem ou não ter o atributo MORF:

- Na categoria LOCAL, o tipo VIRTUAL.
- Na categoria TEMPO, os tipos DATA e PERIODO.

Uma série de exemplos de aplicação são apresentados posteriormente para clarificar em que situações ocorrem estas excepções.

17.1.1 Género (morfológico)

Consideramos que o género de uma EM pode ter três valores:

- M:** EM com género masculino.
- F:** EM com género feminino.
- ?:** Para os casos em que o género é indefinido.

17.1.2 Número

Consideramos que o número de uma EM pode ter três valores:

- S:** EM no singular.
- P:** EM no plural.
- ?:** Para os casos em que o número é indefinido.

17.1.3 Exemplos de não atribuição de MORF na categoria LOCAL

Em alguns casos particulares do tipo VIRTUAL, o atributo MORF foi omitido, devido ao facto de não ser possível avaliar morfologicamente números de telefone.

Certo: <LOCAL TIPO="VIRTUAL">(48) 281 9595</LOCAL>

Os casos que possuam a etiqueta MORF são, pelo contrário, geralmente casos em que a entidade é de outro tipo básico, mas é empregue no contexto na acepção de LOCAL.

Certo: Como capturar da <LOCAL TIPO="VIRTUAL" MORF="F,S">Internet</LOCAL>...

Certo: uma ordem do governo local publicada na "<LOCAL TIPO="VIRTUAL" MORF="F,S">Gazeta de Macau</LOCAL>" ordenava...

Certo: E só depois da publicação no '<LOCAL TIPO="VIRTUAL" MORF="M,S">Diário da República</LOCAL>' é que tomou-se conhecimento do traçado.

17.1.4 Exemplos de não atribuição de MORF na categoria TEMPO

Nos tipos PERIODO e DATA há casos distintos em que são aplicados o atributo MORF.

As datas especificadas em termos de anos ou de dias não possuem nunca a etiqueta MORF.

Certo: Este ano de <TEMPO TIPO="PERIODO">1982</TEMPO> deve...

Certo: <TEMPO TIPO="PERIODO">1914-1918</TEMPO>...

Certo: ia ser a <TEMPO TIPO="DATA">17 de Dezembro</TEMPO> porque saiu...

Certo: Em <TEMPO|TEMPO TIPO="DATA|PERIODO">91</TEMPO>, foram angariados...

As classificações que possuem atributo MORF são meses, séculos, e períodos históricos .

Certo: Cinema para o mês de <TEMPO TIPO="PERIODO" MORF="M,S">Maio</TEMPO>.

Certo: Mas já vem do <TEMPO TIPO="DATA" MORF="M,S">século XVI</TEMPO> o feriado.

Certo: os povoadores cristãos da <TEMPO|ACONTECIMENTO TIPO="PERIODO |EFEMERIDE" MORF="F,S">Reconquista</TEMPO|ACONTECIMENTO>.

Certo: Nesta <TEMPO TIPO="PERIODO" MORF="F,S">Primavera</TEMPO>, encontrei-me com os meus amigos.

Certo: está agora previsto para <TEMPO TIPO="DATA" MORF="M,S">Outubro</TEMPO> ou <TEMPO TIPO="DATA" MORF="M,S">Novembro</TEMPO>

17.2 Regras de atribuição de classificação morfológica

Considera-se o contexto e o texto adjacente para determinar o género e o número de uma dada EM, que à partida pode não ter género ou número definido.

Quando nem esse contexto nem o conhecimento lexical dos anotadores permite atribuir valores definidos, usa-se o valor '?', não especificado.

Exemplos:

Certo: O <PESSOA TIPO="INDIVIDUAL" MORF="M,S">João</PESSOA> é um professor.

Certo: A <PESSOA TIPO="INDIVIDUAL" MORF="F,S">João</PESSOA> não veio.

Certo: O apelido <ABSTRACCAO TIPO="NOME" MORF="?,S">João</ABSTRACCAO> é muito raro.

Ou seja, o nome *João* tem diferentes interpretações da sua classificação morfológica, consoante o contexto em que se encontra inserido.

17.2.1 Exemplos na categoria LOCAL

Algumas localidades administrativas são precedidas por artigo, determinando assim o género e número da entidade que designam (*o Porto, a Madeira, o Brasil, a Guarda, o Minho, o Rio Grande do Sul, os Estados Unidos*). Contudo, muitas outras não levam artigo e torna-se mais difícil de atribuir uma classificação morfológica.

Pareceu-nos em alguns casos haver consenso, tal como para *Portugal* (M, S), *Lisboa* (F, S), *Bragança* (F, S), *Brasília* (F, S), *Nova Iorque* (F, S) e *Colónia* (F, S), mas noutros casos apenas pudemos usar '?' no género, tal como em *Chaves, São Paulo* (estado ou cidade), *Castelo Branco, Braga* ou *Madrid*, excepto quando tal é especificado no contexto.

Certo: <LOCAL TIPO=ADMINISTRATIVO MORF="F,S">Leiria</LOCAL> é linda.

Certo: do concelho de <LOCAL TIPO="ADMINISTRATIVO" MORF="?,S">Aregos</LOCAL>.

Certo: todo o noroeste(de <LOCAL TIPO="ADMINISTRATIVO" MORF="?,S">Resende</LOCAL> ao...

Certo: em <LOCAL TIPO="ADMINISTRATIVO" MORF="M,S">Portugal</LOCAL> seria...

Certo: ...aqui em <LOCAL TIPO="ADMINISTRATIVO" MORF="M,S">São Paulo</LOCAL>.

Certo: ...em <LOCAL TIPO="ADMINISTRATIVO" MORF="F,S">Nova Iorque</LOCAL> e saímos...

Certo: ...polícia de <LOCAL TIPO="ADMINISTRATIVO" MORF="F,S">Colónia</LOCAL> foram suspensos...

17.2.2 Exemplos na categoria ORGANIZACAO

Geralmente o número e género de uma organização são definidos pelo número e género da primeira palavra do nome, *Charcutaria Brasil* (F,S), *Armazéns do Chiado* (M,P), *Banco X* (M,S) ou *Caixa Y* (F,S), enquanto empresas internacionais têm geralmente associado o género feminino: *A Coca-Cola, a Benetton, a IBM, a Microsoft, a Sun, a Lotus, a Ferrari, etc.*

Certo: junto do <ORGANIZACAO TIPO="EMPRESA" MORF="M,S">Banco Sotto Mayor</ORGANIZACAO>.

Certo: Uma acção da <ORGANIZACAO TIPO="EMPRESA" MORF="F,S">Cartier</ORGANIZACAO>.

Certo: A acção da <ORGANIZACAO TIPO="EMPRESA" MORF="F,S">Portugal Telecom</ORGANIZACAO> resultou...

Certo: Esta página tem o apoio da <ORGANIZACAO TIPO="EMPRESA" MORF="F,S">IP</ORGANIZACAO>.

17.2.3 Exemplos na categoria PESSOA

No caso de GRUPOMEMBRO, ou seja, grupos de pessoas, o número é geralmente plural, e o género depende do sexo dos membros. *As Doce, os ABBA, os Xutos e Pontapés, os Beatles, as Spice Girls, os GNR...*

Certo: os <PESSOA TIPO="GRUPOMEMBRO" MORF="M,P">Stones</PESSOA>

Certo: e antes dos <PESSOA TIPO="GRUPOMEMBRO" MORF="M,P">R.E.M.</PESSOA>

Certo: <PESSOA TIPO="GRUPOMEMBRO" MORF="M,P">Peruanos</PESSOA> com diamantes falsos.

Certo: depois os <PESSOA TIPO="GRUPOMEMBRO" MORF="M,P">Mouros</PESSOA> que

lhe deram o nome...

Certo: ...dez minutos o <PESSOA TIPO="GRUPOMEMBRO" MORF="M,S">Bastia
</PESSOA>assegurou a presença na final...

17.2.4 Exemplos na categoria ACONTECIMENTO

No caso do tipo EVENTO, os acontecimentos desportivos que tenham duas equipas, o número é singular, e o género é masculino, visto que correspondem a um jogo.

Certo: seguintes jogos: <ACONTECIMENTO TIPO="EVENTO" MORF="M,S">
Penafiel-Rio Ave</ACONTECIMENTO>

Certo: e o <ACONTECIMENTO TIPO="EVENTO" MORF="M,S">
Nacional-Académica</ACONTECIMENTO>

17.2.5 Exemplos na categoria ABSTRACCAO

No caso do tipo DISCIPLINA, a maior parte das EM que se refiram a disciplinas na área da educação tem género feminino, o número pode variar consoante o primeiro átomo.

Certo: e <ABSTRACCAO TIPO="DISCIPLINA" MORF="F,S">Filosofia</ABSTRACCAO>
em todas as universidades.

Certo: <ABSTRACCAO TIPO= DISCIPLINA MORF= F,S >Ciência da Informação
</ABSTRACCAO>.

Certo: futuros professores de <ABSTRACCAO TIPO="DISCIPLINA" MORF="F,S">
Educação Física</ABSTRACCAO>.

Certo: As <ABSTRACCAO TIPO="DISCIPLINA" MORF="F,P">TI</ABSTRACCAO> são
uma ferramenta...

Já em relação a desportos, o género é em geral masculino, embora haja alguns que, por serem originários de palavras portuguesas femininas, mantêm o género, tal como *Vela* ou *Luta livre*.

Certo: Página do time de <ABSTRACCAO TIPO="DISCIPLINA" MORF="M,S">
Handebol</ABSTRACCAO>