

Segundo HAREM

Modelo Geral, Novidades e Avaliação

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

Encontro do Segundo HAREM
Universidade de Aveiro, 7 de Setembro de 2008

Desafios

- **HAREM clássico** (seguindo genericamente o modelo de classificação proposto no Primeiro HAREM, mas refinando algumas distinções, quer ao nível da segmentação das EM quer da sua classificação)
- Pista para identificação e normalização de **expressões temporais** (Hagège et al., 2008)
- Pista ReReLEM, para a identificação de **relações semânticas entre EM**

Calendário do HAREM (2007)

Outubro: 31

Novembro: 30

Primeira chamada para o Segundo HAREM

Versão preliminar das directivas do Segundo HAREM

Data-limite para discussão sobre os moldes do Segundo HAREM

Calendário do HAREM (2008)

Janeiro: 31

Abril: 30

Junho: 30

Primeira versão da mini-CD do TEMPO e a versão final CD

Primeira versão da mini-CD do ReReLEM

Versão final da mini-CD do TEMPO

Resultados preliminares do HAREM clássico

Resultados preliminares da pista do TEMPO

Directivas finais do material de

Colecção HAREM

Primeira versão da Colecção Dourada

Resultados da pista ReReLEM

Relatórios de desempenho

Versão final da mini-CD do ReReLEM

Encontro tipo Segundo HAREM

Caracterização do HAREM

EM em contexto

Regressou então a <EM ID="hub-67792-106" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">Portugal, onde iniciou meteórica carreira na experimentação de novas formas de expressão

O acordo político quanto à revisão foi obtido durante a <EM ID="aa46996-48" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Presidência Alemã, tendo cabido a <EM ID="aa46996-54" CATEG="ORGANIZACAO" TIPO="ADMINISTRACAO">Portugal concluir o processo de revisão.

Este debate passou completamente ao lado de <EM ID="H2-dftr765-8" CATEG="PESSOA" TIPO="POVO">Portugal

«o problema do <EM ID="hub-24360-322" CATEG="ORGANIZACAO" ABSTRACAO" TIPO="INSTITUICAO" DISCIPLINA">PSD é começar a ter só um <EM ID="hub-24360-323" CATEG="ABSTRACAO" TIPO="IDEIA">Portugal ou dois dentro de si»

<EM ID="Ex-1GM" CATEG="PESSOA" TIPO="GRUPOMEMBRO">Portugal perdeu com a <EM ID="Ex-2GM" CATEG="PESSOA" TIPO="GRUPOMEMBRO">Suíça por <EM ID="Ex-3GM" CATEG="VALOR" TIPO="CLASSIFICACAO">2-0

Caracterização do HAREM

EM vagas

Pela mão do <EM ID="aa66435-107" CATEG="PESSOA" TIPO="INDIVIDUAL">ministro Freitas do Amaral, e sem necessidade alguma, <EM ID="aa66435-108" CATEG="ORGANIZACAO" TIPO="ADMINISTRACAO" POVO">Portugal foi enxovalhado, coberto de vergonha e de cobardia, por um dos mais tristes textos políticos que já alguém escreveu.

<EM ID="hub-28874-31" CATEG="VALOR" TIPO="QUANTIDADE">Mais de 32 mil pessoas poderiam morrer se uma pandemia de gripe humana de origem aviária atingisse <EM ID="hub-28874-32" CATEG="PESSOA" LOCAL" TIPO="POVO" HUMANO" SUBTIPO="PAIS">Portugal

os dois reinos católicos, <EM ID="aa66435-51" CATEG="PESSOA" ORGANIZACAO" TIPO="GRUPOIND" ADMINISTRACAO">Portugal e <EM ID="aa66435-52" CATEG="PESSOA" ORGANIZACAO" TIPO="GRUPOIND" ADMINISTRACAO">Espanha, partiram à conquista do mundo e tornaram-se <ALT"><EM ID="aa66435-53" CATEG="OUTRO">Impérios | Impérios </ALT"> marítimos do <EM ID="aa66435-54" CATEG="LOCAL" LOCAL" TIPO="FISICO" HUMANO" SUBTIPO="REGIAO" DIVISAO">Novo Mundo

Caracterização do HAREM

■ Análises alternativas (ALT)

➤ Ambiguidade estrutural:

aproximava a <ALT> <EM ID="H2-dftre765-106" CATEG="ABSTRACCAO" TIPO="DISCIPLINA">Igreja de Inglaterra | <EM ID="H2-dftre765-106-aa" CATEG="ABSTRACCAO" TIPO="DISCIPLINA">Igreja de <EM ID="H2-dftre765-16" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">Inglaterra </ALT> do calvinismo.

➤ Possibilidade de uma expressão poder ser reconhecida como EM, ou não:

<EM ID="aa66435-51" CATEG="PESSOA|ORGANIZACAO" TIPO="GRUPOIND|ADMINISTRACAO">Portugal e <EM ID="aa66435-52" CATEG="PESSOA|ORGANIZACAO" TIPO="GRUPOIND|ADMINISTRACAO">Espanha, partiram à conquista do mundo e tornaram-se <ALT> <EM ID="aa66435-53" CATEG="OUTRO">Impérios | Impérios </ALT> marítimos.

7

Novidades no Segundo HAREM

■ Uso extensivo de ALT para representação sistemática da estrutura interna das EM constituídas por outras EM

<ALT>

<EM ID="aa55968-473" CATEG="PESSOA" TIPO="CARGO">presidente da Câmara de Nova Iorque

| presidente da

<EM ID="aa55968-7" CATEG="ORGANIZACAO" TIPO="ADMINISTRACAO">Câmara de Nova Iorque

| presidente da

<EM ID="aa55968-475aa" CATEG="ORGANIZACAO" TIPO="ADMINISTRACAO">Câmara de <EM ID="aa55968-475" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="DIVISAO">Nova Iorque

</ALT>

8

Novidades no Segundo HAREM

■ ALT sistematicamente aplicados

CATEG	ESTRUTURA	EXEMPLO
PESSOA	PESSOA de LOCAL PESSOA de ORGANIZACAO PESSOA de ORG LOCAL PESSOA de PESSOA	Santo António de Lisboa Presidente do Parlamento Europeu Presidente dos Estados Unidos FCP de José Mourinho
ORGANIZACAO	ORG de LOCAL ORG de ORG	Associação Académica de Coimbra Conselho da UE
LOCAL	LOCAL de LOCAL LOCAL de ORGANIZACAO	Teatro Scala de Milão Centro de Eventos da Fiergs
OBRA	OBRA de PESSOA OBRA LOCAL de LOCAL	Sagrada Família de Gaudí Mosteiro de Santa Clara-a-Velha
ACONTECIMENTO	ACONT de LOCAL ACONT de TEMPO	Jogos Olímpicos de Barcelona Volta a Espanha de 1985
ABSTRACCAO	ABSTRACCAO de LOCAL	Blues do Delta
COISA	COISA de ORGANIZACAO	iPhone da Apple

9

Novidades no Segundo HAREM

■ Não reconhecimento de certas palavras em maiúsculas como EM

➤ Quando ocorrem no âmbito de uma unidade lexical composta que não constitui, ela própria, uma EM
pastéis de Belém; amêijoas à Bulhão Pato

➤ Se os restantes constituintes (grafados em minúscula) dessa unidade lexical integrarem a lista das minúsculas permitidas no âmbito desta avaliação, toda a unidade lexical deverá ser reconhecida como EM

<EM ID="ric-19806-172" CATEG="ABSTRACCAO" TIPO="ESTADO">doença de Chagas

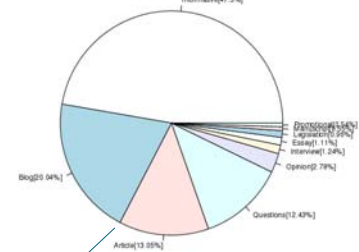
10

CATEGORIA	TIPO	SUBTIPO
ABSTRACCAO	ESTADO, DISCIPLINA, IDEIA, NOME OUTRO	X ESCOLA, OBRA, PLANO
ACONTECIMENTO	EFEMERIDE, EVENTO, ORGANIZADO OUTRO	
COISA	CLASSE, MEMBROCLASSE, OBJECTO SUBSTANCIA, OUTRO	
LOCAL	FISICO HUMANO VIRTUAL OUTRO	AGUACURSO, AGUAMASSA, ILHA, PLANETA, REGIAO, RELEVO, OUTRO CONSTRUCAO, DIVISAO, PAIS, REGIAO, RUA, OUTRO COMSOCIAL, OBRA, SITIO, OUTRO OUTRO
OBRA	ARTE, PLANO, REPRODUZIDA, OUTRO	
ORGANIZACAO	ADMINISTRACAO, EMPRESA, INSTITUICAO OUTRO	X SUB
PESSOA	CARGO, GRUPOCARGO, GRUPOIND, GRUPOMEMBRO, INDIVIDUAL, MEMBRO, POVO, OUTRO	
TEMPO	DURACAO, FREQUENCIA, GENERICO TEMPO, CALEND OUTRO	X PERIODO, CICLICO TRO
VALOR	CLASSIFICACAO, MOEDA, QUANTIDADE, OUTRO	
OUTRO		X VARIADO

13

Colecção HAREM

DOCS: 1040
Parágrafos: 15737
Palavras: 670610



Distribuição dos documentos por género na CD

14

Programas de avaliação do HAREM clássico: entradas

CDHAREM.xml

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<CDHAREM version="SegundoHarem">
...
<DOC DOCID="Xyz">
  O <ALT><EM ID="Xyz_60" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França de 2009</EM> <EM ID="Xyz_61" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França</EM> <EM ID="Xyz_62" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <EM ID="Xyz_63" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour</EM> <EM ID="Xyz_64" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">França</EM> <EM ID="Xyz_65" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <ALT> vai começar no <EM ID="Xyz_584" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">Mónaco</EM> com um contra-relógio de <EM ID="Xyz_585" CATEG="VALOR" TIPO="QUANTIDADE">15 quilómetros</EM>, numa etapa que também terminará no principado monégaco. O início do contra-relógio terá lugar precisamente no mesmo local do local de partida das provas de Fórmula 1. "O prestígio do principado, a filosofia que tem para o desporto e grandes eventos e a localização geográfica" foram apontados por <EM ID="Xyz_587" CATEG="PESSOA" TIPO="INDIVIDUAL">Christian Prudhomme</EM> como justificação para a decisão.
</DOC>
</CDHAREM>
```

participacao.xml

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<CDHAREM version="OD">
...
<DOC DOCID="Xyz">
  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> vai começar no <EM ID="Xyz_3" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">Mónaco</EM> com um contra-relógio de <EM ID="Xyz_4" CATEG="VALOR" TIPO="QUANTIDADE">15 quilómetros</EM>, numa etapa que também terminará no principado monégaco. O início do contra-relógio terá lugar precisamente no mesmo local do local de partida das provas de Fórmula 1. "O prestígio do principado, a filosofia que tem para o desporto e grandes eventos e a localização geográfica" foram apontados por <EM ID="Xyz_5" CATEG="PESSOA" TIPO="INDIVIDUAL">Christian Prudhomme</EM> como justificação para a decisão.
</DOC>
</CDHAREM>
```

Programas de avaliação do HAREM clássico: Alinhador

CDHAREM.xml

participacao.xml

Alinhador

EMs na CD

EMs na participação

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<CDHAREM version="SegundoHarem">
...
<DOC DOCID="Xyz">
  O <ALT><EM ID="Xyz_60" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França de 2009</EM> <EM ID="Xyz_61" CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM> <EM ID="Xyz_62" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <EM ID="Xyz_63" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour</EM> <EM ID="Xyz_64" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">França</EM> <EM ID="Xyz_65" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <ALT> vai começar no <EM ID="Xyz_584" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">Mónaco</EM> com um contra-relógio de <EM ID="Xyz_585" CATEG="VALOR" TIPO="QUANTIDADE">15 quilómetros</EM>, numa etapa que também terminará no principado monégaco. O início do contra-relógio terá lugar precisamente no mesmo local do local de partida das provas de Fórmula 1. "O prestígio do principado, a filosofia que tem para o desporto e grandes eventos e a localização geográfica" foram apontados por <EM ID="Xyz_587" CATEG="PESSOA" TIPO="INDIVIDUAL">Christian Prudhomme</EM> como justificação para a decisão.
</DOC>
</CDHAREM>
```

Programas de avaliação do HAREM clássico: Avaliador de alinhamentos

CDHAREM.xml

participacao.xml

Avaliador de alinhamentos

Alinhamentos classificados como: Correctos, Espúrios, Em Falta ou Parcialmente Correctos

Alinhador

EMs parcialmente removidas na próxima fase

EM em falta

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<CDHAREM version="SegundoHarem">
...
<DOC DOCID="Xyz">
  O <ALT><EM ID="Xyz_60" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França de 2009</EM> <EM ID="Xyz_61" CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM> <EM ID="Xyz_62" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <EM ID="Xyz_63" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour</EM> <EM ID="Xyz_64" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">França</EM> <EM ID="Xyz_65" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <ALT> vai começar no <EM ID="Xyz_584" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">Mónaco</EM> com um contra-relógio de <EM ID="Xyz_585" CATEG="VALOR" TIPO="QUANTIDADE">15 quilómetros</EM>, numa etapa que também terminará no principado monégaco. O início do contra-relógio terá lugar precisamente no mesmo local do local de partida das provas de Fórmula 1. "O prestígio do principado, a filosofia que tem para o desporto e grandes eventos e a localização geográfica" foram apontados por <EM ID="Xyz_587" CATEG="PESSOA" TIPO="INDIVIDUAL">Christian Prudhomme</EM> como justificação para a decisão.
</DOC>
</CDHAREM>
```

Tipos de cenário

- Cenário total:** elenco completo de categorias, tipos e subtipos propostos para o Segundo HAREM.
- Cenário selectivo de avaliação:** sub-conjunto de categorias, tipos e subtipos do cenário total que é alvo de avaliação.
- Cenário selectivo de participação:** sub-conjunto de categorias, tipos e subtipos do cenário total que cada participante se propôs a utilizar.

Programas de avaliação do HAREM clássico: véus HAREM

CDHAREM.xml

participacao.xml

Véus (HAREM)

Avaliador de alinhamentos

Alinhamentos filtrados de acordo com o cenário pretendido

Dentro desta alternativa, a EM é espúria

Remoção de alinhamentos parcialmente correctos

Alinhador

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<CDHAREM version="SegundoHarem">
...
<DOC DOCID="Xyz">
  O <ALT><EM ID="Xyz_60" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França de 2009</EM> <EM ID="Xyz_61" CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM> <EM ID="Xyz_62" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <EM ID="Xyz_63" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour</EM> <EM ID="Xyz_64" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">França</EM> <EM ID="Xyz_65" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <ALT> vai começar no <EM ID="Xyz_584" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">Mónaco</EM> com um contra-relógio de <EM ID="Xyz_585" CATEG="VALOR" TIPO="QUANTIDADE">15 quilómetros</EM>, numa etapa que também terminará no principado monégaco. O início do contra-relógio terá lugar precisamente no mesmo local do local de partida das provas de Fórmula 1. "O prestígio do principado, a filosofia que tem para o desporto e grandes eventos e a localização geográfica" foram apontados por <EM ID="Xyz_587" CATEG="PESSOA" TIPO="INDIVIDUAL">Christian Prudhomme</EM> como justificação para a decisão.
</DOC>
</CDHAREM>
```

Programas de avaliação do HAREM clássico: Organizador de ALTs

CDHAREM.xml

participacao.xml

Organizador de ALTs

Véus (HAREM)

Avaliador de alinhamentos

Alinhador

A EM espúria desapareceu porque alinhava correctamente com a alternativa seguinte

Peso

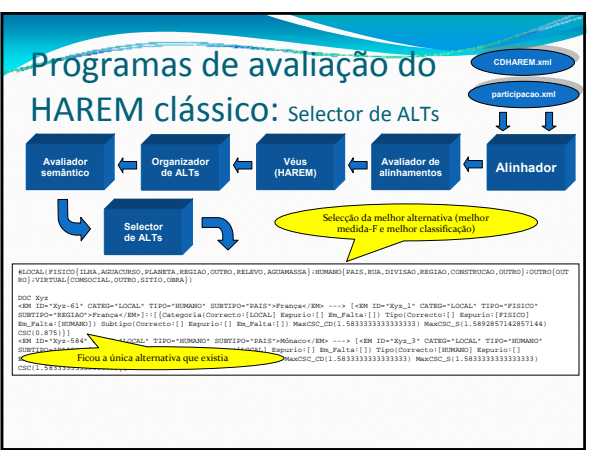
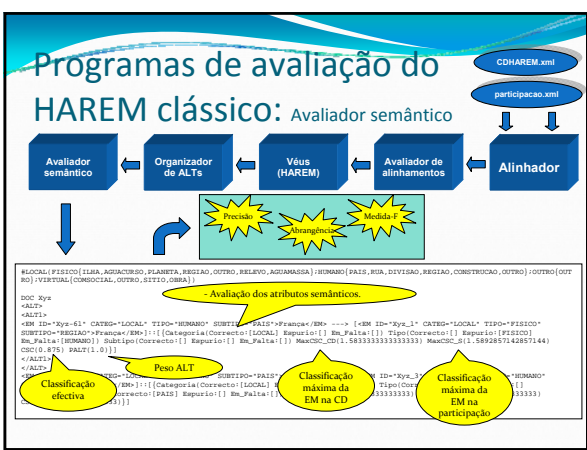
Remoção de alternativas vazias/duplicadas, surgidas depois da aplicação de véus

Remoção de EMs alinhadas mais do que uma vez dentro do mesmo ALT

Atribuição do peso/total de ALT a cada EM dentro de um ALT

Alinhador

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<CDHAREM version="SegundoHarem">
...
<DOC DOCID="Xyz">
  O <ALT><EM ID="Xyz_60" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour de França de 2009</EM> <EM ID="Xyz_61" CATEG="LOCAL" TIPO="FISICO" SUBTIPO="REGIAO">França</EM> <EM ID="Xyz_62" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <EM ID="Xyz_63" CATEG="ACONTECIMENTO" TIPO="ORGANIZADO">Tour</EM> <EM ID="Xyz_64" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">França</EM> <EM ID="Xyz_65" CATEG="TEMPO" TIPO="TEMPO_CALEND" SUBTIPO="DATA">de 2009</EM> <ALT> vai começar no <EM ID="Xyz_584" CATEG="LOCAL" TIPO="HUMANO" SUBTIPO="PAIS">Mónaco</EM> com um contra-relógio de <EM ID="Xyz_585" CATEG="VALOR" TIPO="QUANTIDADE">15 quilómetros</EM>, numa etapa que também terminará no principado monégaco. O início do contra-relógio terá lugar precisamente no mesmo local do local de partida das provas de Fórmula 1. "O prestígio do principado, a filosofia que tem para o desporto e grandes eventos e a localização geográfica" foram apontados por <EM ID="Xyz_587" CATEG="PESSOA" TIPO="INDIVIDUAL">Christian Prudhomme</EM> como justificação para a decisão.
</DOC>
</CDHAREM>
```



Participação no Segundo HAREM

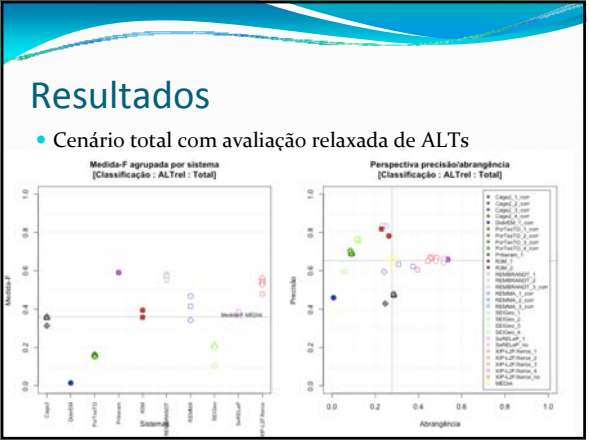
- Manifestações de interesse: 22
- Sistemas inscritos: 16
- Sistemas participantes: 10
- Saídas: 27 (máx. 4 por participante)

Cenários selectivos de participação

Sistema	CEN	PES	ORG	LOC	OBR	ACO	ABS	COI	TEM	VAL
Cagea	Sel2	CAT	CAT	F + H					CAT	
DobrEM	Pes									
PorTexTO	Temp									
Priberam	Tot									
R3M	Sel3									
REMBRANDT	Tot									
REMMMA	Sel4			C/T					C/T	
SEI-Geo	Sel5			F + H						
SeRELeP	Tot									
XIP/LaF/XEROX	Sel6								NORM	

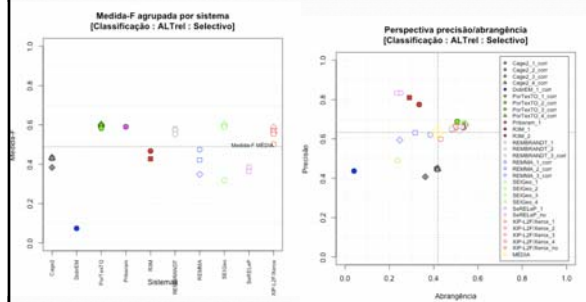
■ Identificação
■ Classificação

CAT – apenas CATEGORIA
 C/T – apenas CATEGORIA e TIPO
 F+H – apenas locais FÍSICOS e HUMANOS
 NORM – normalização das expressões temporais

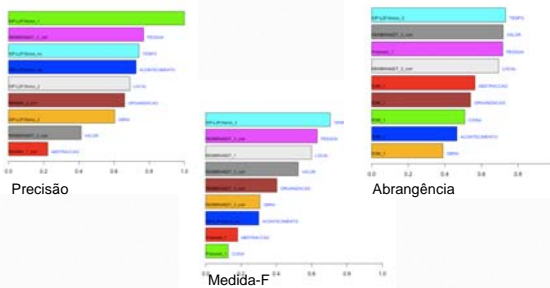


Resultados

- Cenário próprio com avaliação relaxada de ALTs



Avaliação por categorias (melhores sistemas)



34

Agradecimentos

- A Linguateca e o HAREM são financiados através do contrato nº 339/1.3/C/NAC, financiado pelo governo português e pela União Europeia, e executado pela FCCN.

POS CONHECIMENTO

União Europeia
FEDER

UMIC
Unidade de Informação e Multimédia

FCCN
Fundação para a Computação Científica Nacional

35