

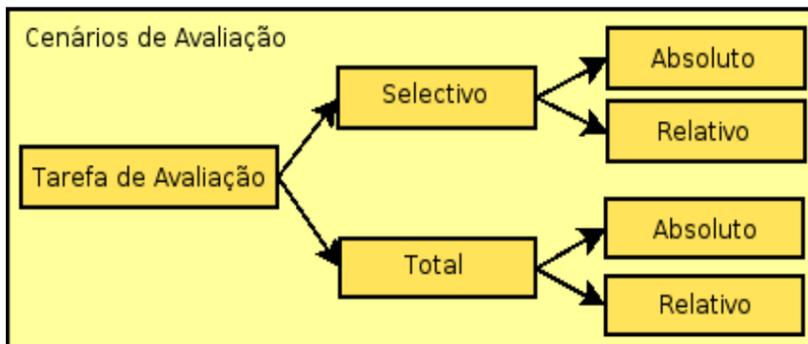
HAREM - Avaliação conjunta de sistemas de reconhecimento de entidades mencionadas

Linguatca - <http://www.linguatca.pt>

Diana Santos, Nuno Seco, Nuno Cardoso, Rui Vilela

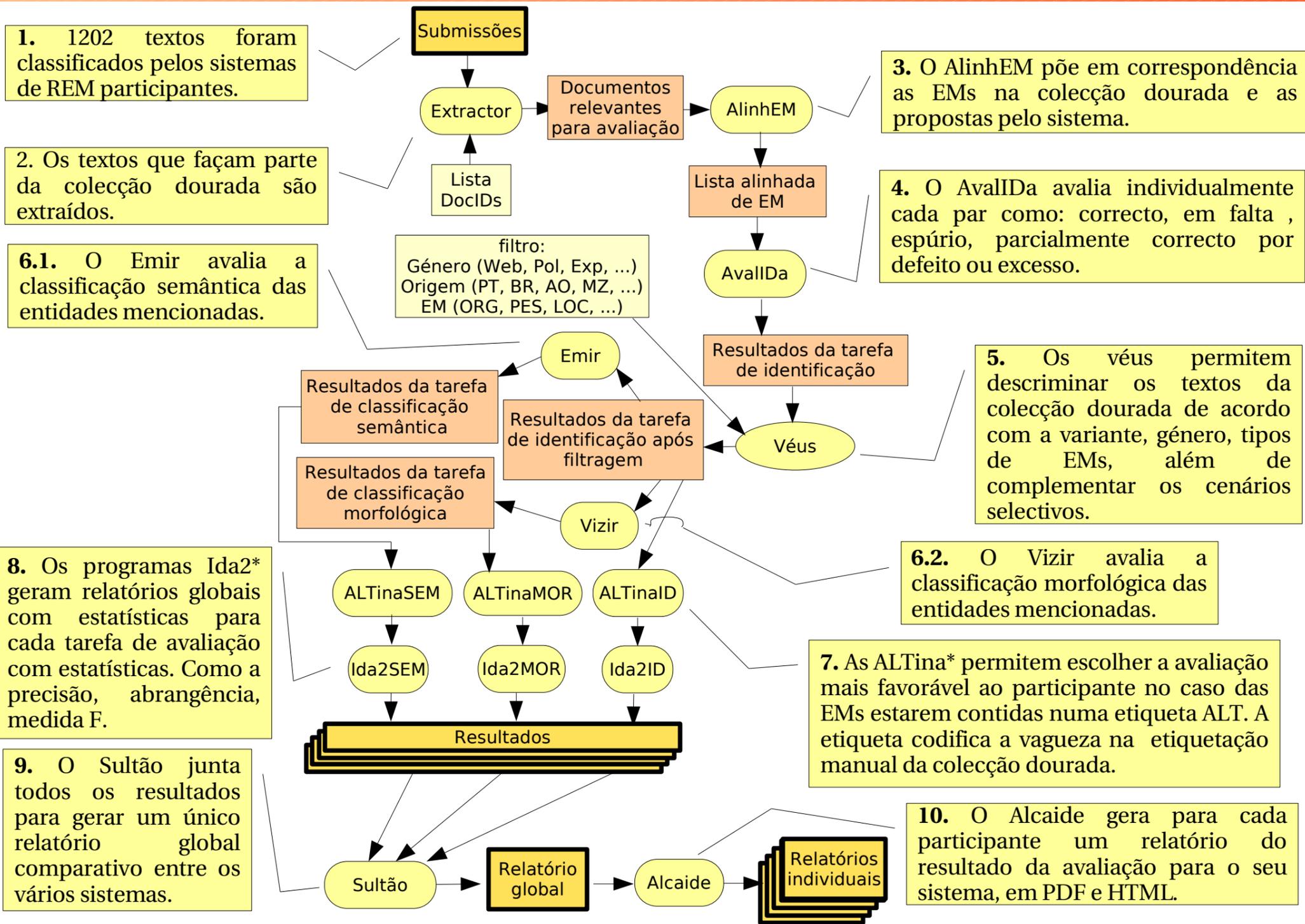
- HAREM foi a primeira avaliação conjunta para o reconhecimento de entidades mencionadas em português.
- Os programas de avaliação foram implementados após o desenvolvimento de um novo modelo de avaliação.

- 129 do HAREM + 128 do miniHAREM textos manualmente classificados formam a colecção dourada, com 5270 + 3851 EMs.
- Os textos escolhidos são originários de vários países lusófonos.
- Vários géneros de texto foram usados.



Um conjunto de métodos e medidas são usados para classificar a avaliação dos resultados dos sistemas de REM. A avaliação é dividida em 3 tarefas:

- Identificação de entidades mencionadas
- Classificação semântica
- Classificação morfológica



1. 1202 textos foram classificados pelos sistemas de REM participantes.

2. Os textos que fazem parte da colecção dourada são extraídos.

6.1. O Emir avalia a classificação semântica das entidades mencionadas.

3. O AlinhEM põe em correspondência as EMs na colecção dourada e as propostas pelo sistema.

4. O AvalIDa avalia individualmente cada par como: correcto, em falta, espúrio, parcialmente correcto por defeito ou excesso.

5. Os véus permitem discriminar os textos da colecção dourada de acordo com a variante, género, tipos de EMs, além de complementar os cenários selectivos.

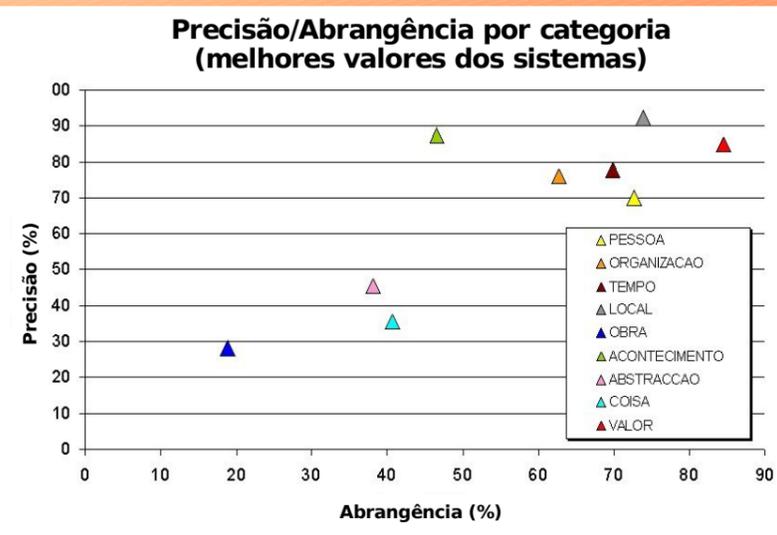
6.2. O Vizir avalia a classificação morfológica das entidades mencionadas.

8. Os programas Ida2* geram relatórios globais com estatísticas para cada tarefa de avaliação com estatísticas. Como a precisão, abrangência, medida F.

7. As ALTina* permitem escolher a avaliação mais favorável ao participante no caso das EMs estarem contidas numa etiqueta ALT. A etiqueta codifica a vagueza na etiquetagem manual da colecção dourada.

9. O Sultão junta todos os resultados para gerar um único relatório global comparativo entre os vários sistemas.

10. O Alcaide gera para cada participante um relatório do resultado da avaliação para o seu sistema, em PDF e HTML.



- A arquitectura de avaliação do HAREM adiciona grande flexibilidade à avaliação de EMs.
- Os programas desenvolvidos são públicos e de código aberto, assim como a colecção dourada, fornecendo um ponto de partida a qualquer investigador que pretenda avaliar sistemas de EMs para o português.
- Confirmam-se as diferenças em EMs dos géneros textuais.
- Várias características inovadoras.