

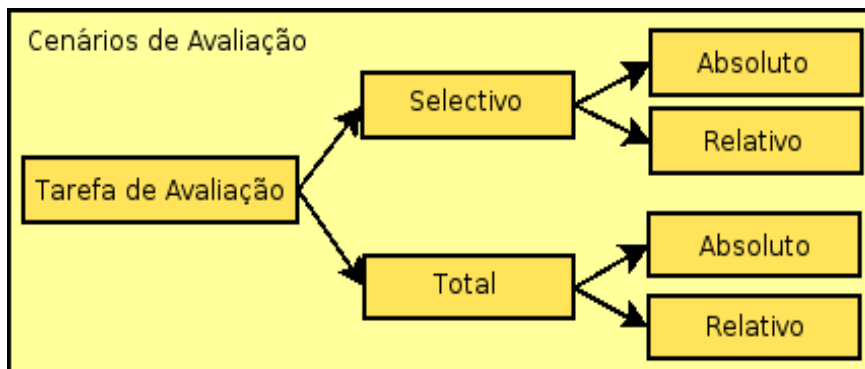
# HAREM - Avaliação conjunta de sistemas de reconhecimento de entidades mencionadas

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

Diana Santos, Nuno Seco, Nuno Cardoso, Rui Vilela

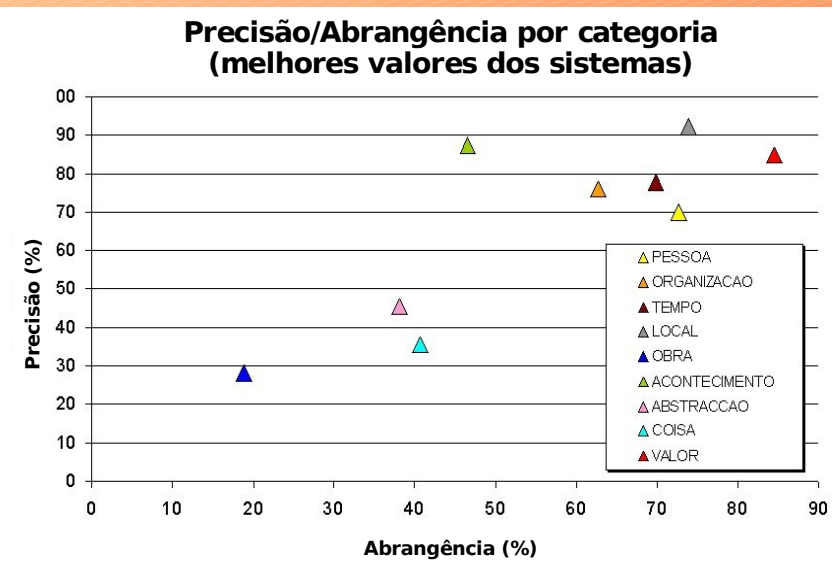
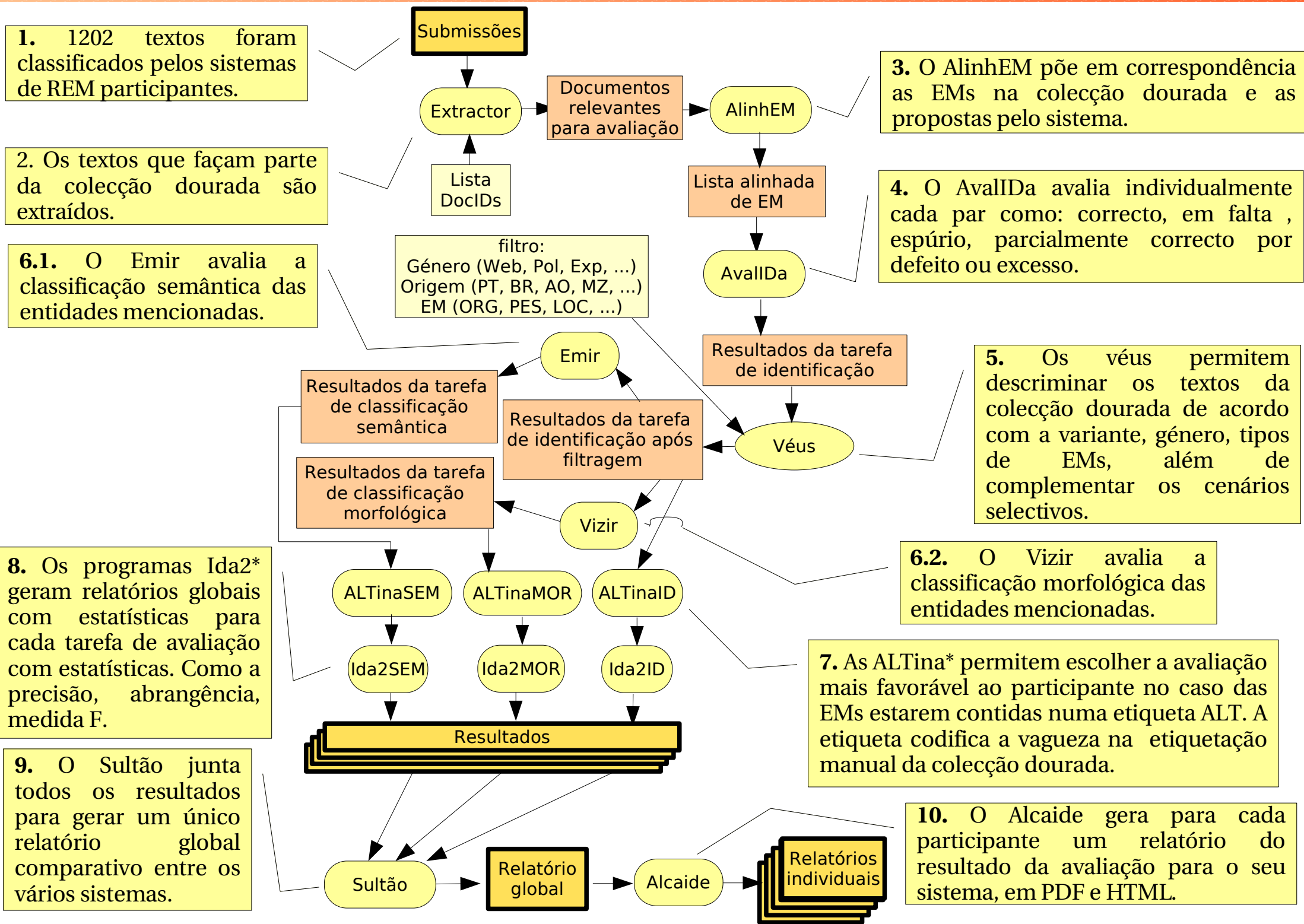
- HAREM foi a primeira avaliação conjunta para o reconhecimento de entidades mencionadas em português.
- Os programas de avaliação foram implementados após o desenvolvimento de um novo modelo de avaliação.

- 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



- 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.