

Identificação de Personagens: **Outras métricas de avaliação**

Prof. Roberto Willrich (INE-UFSC)



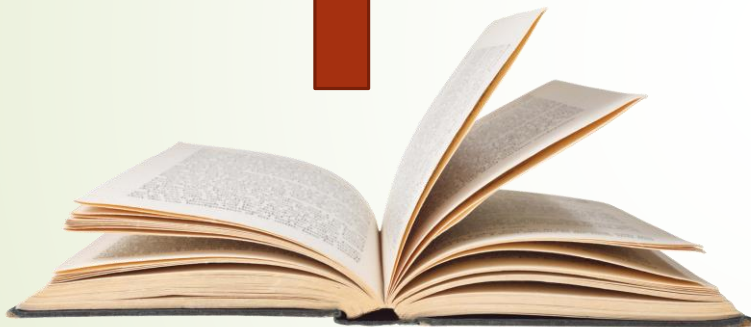
Plano da Apresentação

- 1. DIP: Classificando o Problema DIP
- 2. DIP: Principais Etapas de uma técnica
- 3. Resultado esperado das técnicas concorrentes
- 4. Métricas de avaliação utilizada em trabalhos relacionados
- 5. Considerações finais

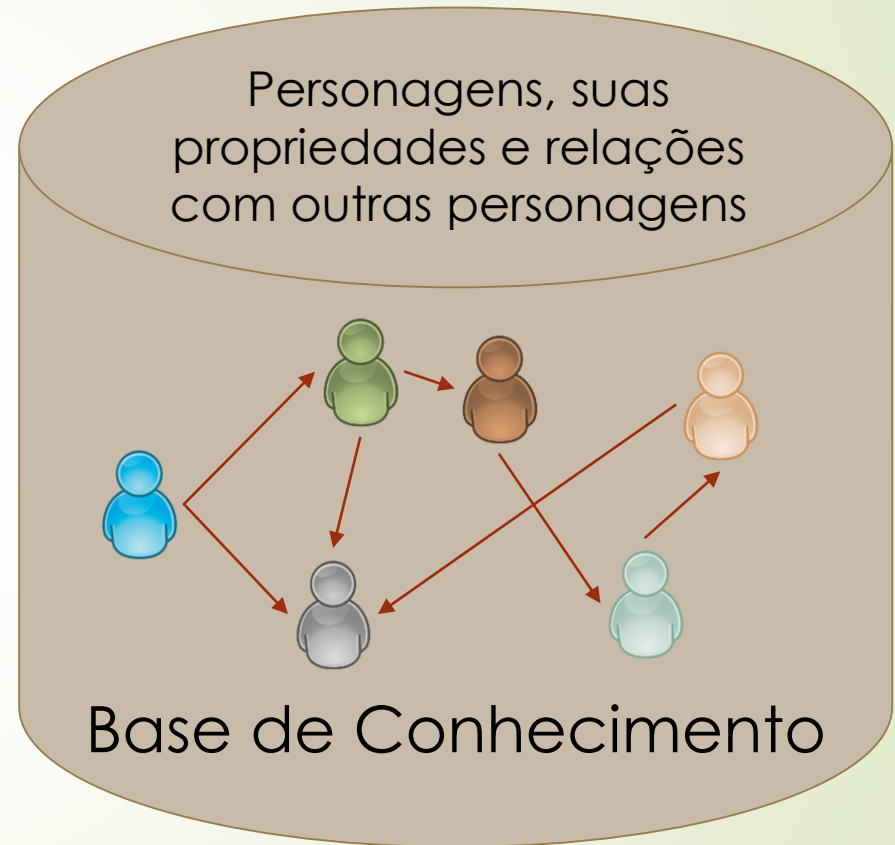


1. DIP: Classificando o problema

Extração do Conhecimento



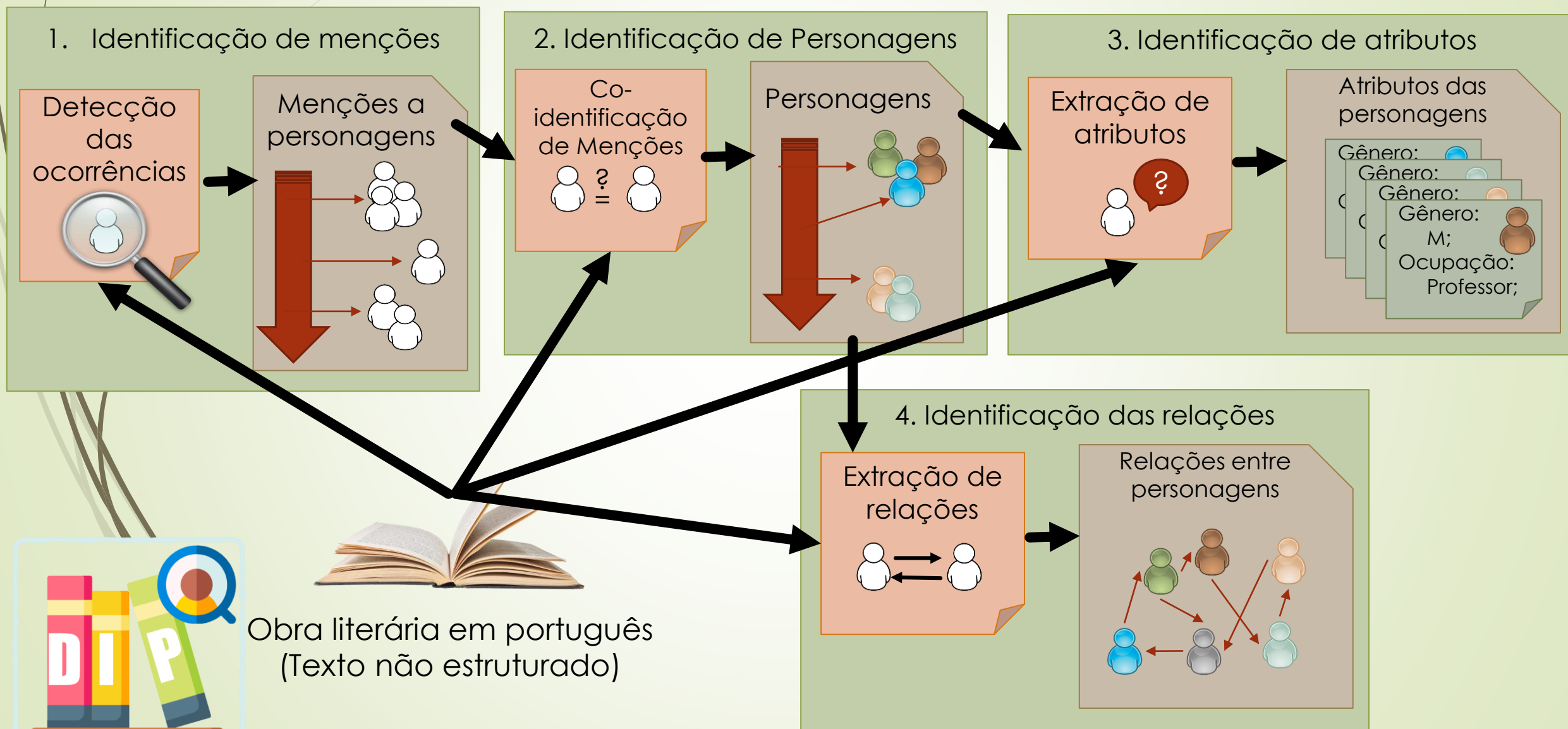
Obra literária em português
(Texto não estruturado)



Como avaliar o conhecimento extraído?



2. DIP: Principais etapas



2. DIP: Resultado esperado das técnicas

➤ Para cada obra, a técnica deve gerar arquivos com dados sobre:

➤ Personagens

➤ Identificador único

➤ Nome(s) de superfície (menções)

➤ Propriedades

➤ Relações entre personagens



Pode ser representado como um Grafo de Conhecimento ou **rede de personagens**



2. Avaliação das Técnicas

- Confrontando resultado do método com obras da coleção dourada

Como avaliar o desempenho/similaridade?



Resultado da técnica candidata



Padrão Ouro



3. DIP: Métodos de avaliação dos Trabalhos relacionados

► Precisão, Revogação e Medida-F em **Recuperação de Informações**

► **Precisão:** Fração das personagens/características/relações identificadas que são corretas

► $P = \frac{\text{número de personagens/características/relações corretamente identificadas}}{\text{número de todas as personagens/características/relações extraídas}}$

► **Revocação** (Abrangência): fração das personagens/características/relações corretas que foram identificadas

► $R = \frac{\text{número de personagens/características/relações corretamente identificadas}}{\text{número de todas as personagens/características/relações da obra}}$

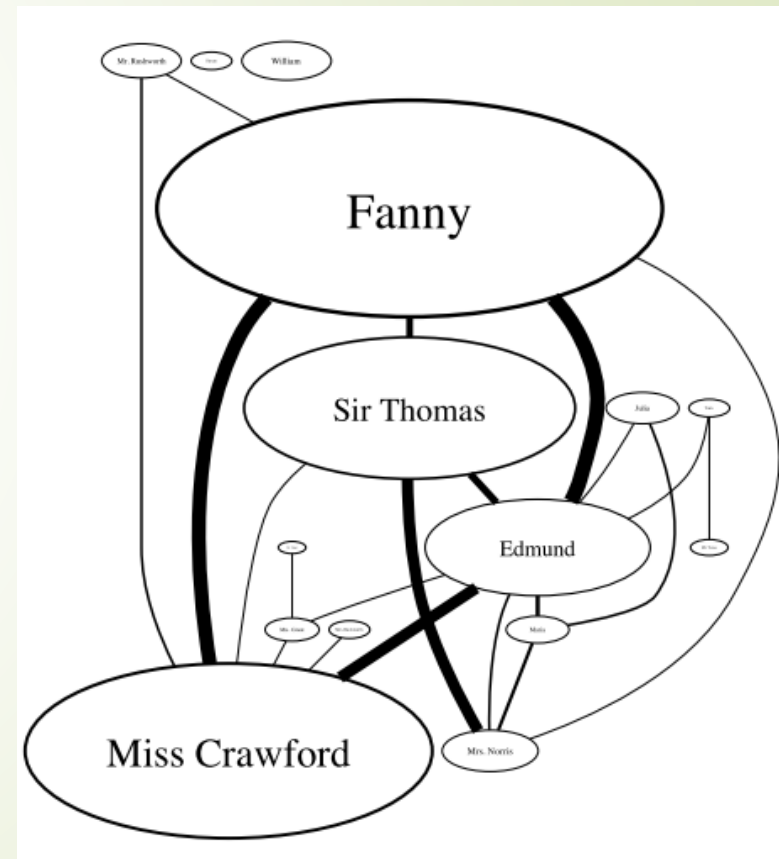
► **Medida-F:** média harmônica entre precisão e revocação

► $F = \frac{2 * P * R}{(P + R)}$



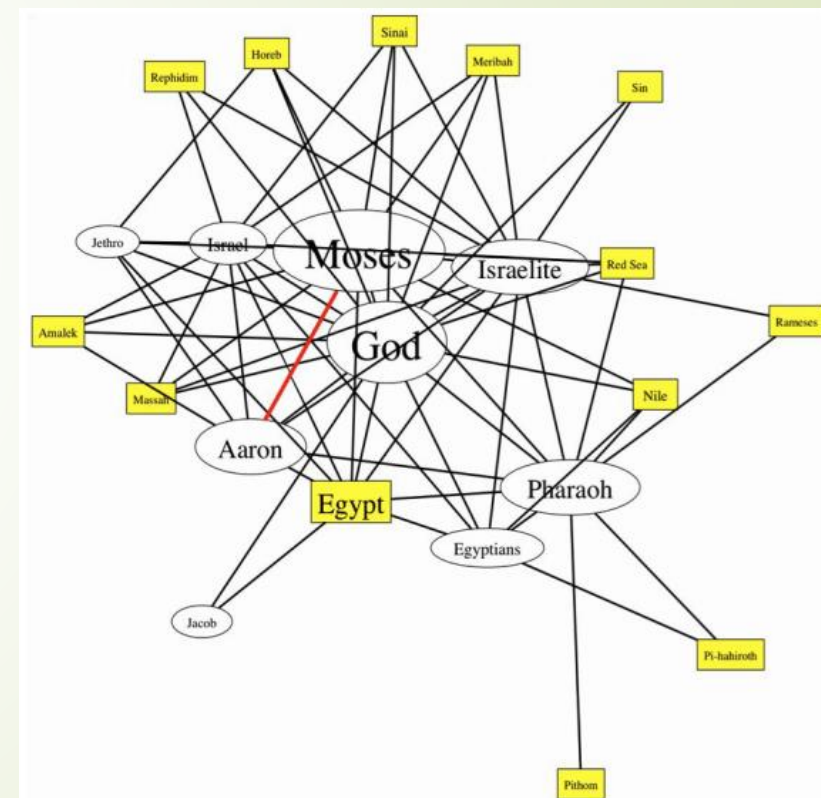
3. DIP: Métodos de avaliação dos Trabalhos relacionados

- ▶ Elson, David K., Kathleen McKeown, and Nicholas J. Dames. "Extracting social networks from literary fiction." (2010).
 - ▶ Objetivo é gerar uma rede conversacional (interações sociais entre personagens):
 - ▶ nodos ponderados de acordo com o nível de interação
 - ▶ arestas ponderadas representando o montante de conversação
 - ▶ Utiliza precisão, revocação e medida-F de arestas personagens-personagens



3. DIP: Métodos de avaliação dos Trabalhos relacionados

- ▶ Lee, John SY, and Chak Yan Yeung. "Extracting networks of people and places from literary texts." *26th Pacific Asia Conf. on Language, Information, and Computation*. 2012.
- ▶ Objetivo é gerar uma rede de personagens e lugares
 - ▶ Arestas entre personagens indicam existência de interação social,
 - ▶ Arestas entre personagens e lugares indicam localização física
- ▶ Utiliza Precisão, revocação e medida-F de arestas personagens-personagens e personagens-lugares



3. DIP: Métodos de avaliação dos Trabalhos relacionados

- Agarwal, Apoorv, Anup Kotalwar, and Owen Rambow. "Automatic extraction of social networks from literary text: A case study on alice in wonderland." Sixth Int. Joint Conf. on Natural Language Processing. 2013.
 - Objetivo: construção da rede de personagens usando NLP
 - Avalia a rede de personagens
 - Usando precisão, revocação e medida-F
 - usando métricas de análise de redes sociais
 - **Centralidade de grau** (grau de comunicação de um personagem) e **de intermediação** (quantifica o número de vezes que um personagem age como ponte ao longo do caminho mais curto entre dois outros personagens), **densidade do grafo**, **componentes conectados**, **triads**
 - Compara métricas do resultado com a coleção dourada usando distância euclidiana dos vetores



3. DIP: Métodos de avaliação dos Trabalhos relacionados

- ▶ Łajewska, Weronika, and Anna Wróblewska. "Protagonists' Tagger in Literary Domain--New Datasets and a Method for Person Entity Linkage." *arXiv preprint arXiv:2110.01349* (2021).
 - ▶ Identificação de Pessoas em obras (personagens e outros nomes citados)
 - ▶ Utiliza método automatizado, baseada em similaridade, para mapear personagens da coleção dourada e do teste.
 - ▶ Utiliza precisão, revocação e medida-F para avaliar a identificação de personagens



3. DIP: Métodos de avaliação dos Trabalhos relacionados

- Schmidt, David, et al. "The FairyNet Corpus-Character Networks for German Fairy Tales." Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature. 2021.
- Corpus com 40 obras: Rede de personagens com: categoria semântica (humano, criatura, genérico), tipo de personagem, idade, arestas com sentimento entre personagens, diferentes tipos de relações familiares (16) e outras relações sociais (13)
- Avalia utilidade do corpus em ferramentas de IP. Formas de avaliação:
 - Precisão, Cobertura e medida-F para categorização semântica das personagens,
 - **Co-identificação de personagens**: avaliada usando precisão, cobertura e medida-f considerando métodos de avaliação de correferência
 - **Medida de concordância** (índice Kappa de Cohen) para avaliar idade, tipo, sentimento, tipo de relação e de sentimento baseada em concordância entre classificadores
 - **Métricas relacionados a estrutura da rede** (grafo) de personagens: número médio de vizinhos de uma entidade, tamanho dos caminhos mais curtos no grafo, diâmetro, centralidade de intermediação e densidade da rede



4. Métrica de avaliação da resolução de Co-identificação de personagens (correferência)

- Identificação do conjunto de menções que se referem à mesma personagem
- Como avaliar a resolução de co-identificação?

Extraídas

- 1 Alberto da Costa|Alberto
- 2 Caetano Ra|Caetano
- 3 Caetano Ramalho|Caetano
- 4 Casa Real
- 5 D. Genoveva
- 6 D. Genoveva Ramalho
- 7 Doutor Luiz de Sandoval|Doutor Luiz Sandoval|Dr. Luiz Sandoval|Dr. Luiz de Sandoval|Luiz de Sandoval|Doutor Sandoval|Dr. Sandoval|Dr. Sandoval|Luiz
- 8 Etelvina Perdoa-me
- 9 Jose Maria de Brito|Jose Maria
- 10 José Maria de Brito
- 11 João
- 12 Luizinho
- 13 Marquez
- 14 Miguel Angelo
- 15 Sr.a D. Bertha|Bertha|sr.a D. Bertha
- 16 Victor Hugo
- 17 Virginia Bella
- 18 capitão Jorge de Albuquerque|capitão Jorge|Jorge

Coleção Dourada

- 1 José Maria de Brito|José Maria|Jose Maria|Jose Maria de Brito
- 2 Caetano Ramalho|Caetano|Caetano Ra
- 3 Etelvina|sr.a D. Etelvina Ramalho de Brito|Etelvina Ramalho
- 4 D. Genoveva|D. Genoveva Ramalho
- 5 Alberto da Costa|Alberto
- 6 Doutor Luiz de Sandoval|Luiz de Sandoval|Dr. Sandoval|Luiz|Doutor Sandoval|L. de S.|Sandoval|Doutor Luiz Sandoval|Dr. Luiz Sandoval|Dr. Luiz de Sandoval
- 7 Bertha de Sandoval|Bertha|Bertha de Albuquerque|Bertha d'Albuquerque|Sr.a D. Berta|sr.a D. Berta
- 8 Antonio|Antonio de Albuquerque
- 9 Jorge|capitão Jorge|capitão Jorge de Albuquerque
- 10 Philomena
- 11 Virginia Bella
- 12 João
- 13 Luizinho



4. Métrica de avaliação da resolução de Co-identificação de personagens (correferência)

- Situações que podem ocorrer:
 - Relação de menções incompletas:
 - 1. José Maria de Brito ~~José Maria~~
 - Personagem não é co-identificada corretamente:
 - 1. José Maria de Brito
 - 2. José Maria
 - Não distingue duas ou mais personagens:
 - 1. José Maria de Brito | **Philomena** | José Maria



4. Métrica de avaliação da resolução de Co-identificação de personagens (correferência)

- ▶ Métricas clássicas da área (não há consenso de qual seria a melhor):
 - ▶ Métrica desempenho MUC determina método para cálculo de P, R e medida F
 - ▶ B-Cubed resolve problemas do anterior
 - ▶ CEAF_e baseada em entidade (Constrained Entity-Alignment F-Measure)
 - ▶ BLANC metric (BiLateral Assessment of Noun-phrase Coreference)
 - ▶ LEA (Link-based Entity-Aware evaluation metric)
 - ▶ CoNLL-Score: média do MCU, B-Cubed e CEAF_e

Referência: Krug, Markus. *Techniques for the Automatic Extraction of Character Networks in German Historic Novels*. Bayerische Julius-Maximilians-Universitaet Wuerzburg (Germany), 2020.



5. Considerações Finais

- Métricas de Precisão, Revocação e Medida F são as mais usadas no domínio
 - Validando as métricas adotadas no DIP
- Recomendações para futuros DIPs:
 - Métodos de mapeamento entre personagens extraídas e da coleção dourada poderiam ser automatizadas
 - Métodos clássicos de avaliação de co-identificação de personagens deveriam ser utilizados

