

# Pré-Projeto - Sistemas Multiagentes

Natan de Souza Rodrigues - 200051741

Outubro de 2020

## 1 Descrição do projeto

A identificação de autores é fundamental para trabalharmos com dados bibliográficos, porém nomes homônimos e sinônimos atrapalham a identificação dos autores em um repositório bibliográfico, sendo assim necessária a aplicação de uma abordagem de desambiguação de nomes de autores.

Atualmente estou trabalhando na pesquisa de desambiguação de nomes, onde obtive resultados iniciais utilizando uma abordagem multi-estratégica para desambiguação, seguindo estratégias de remoção de acentos, similaridade de *strings* - como coeficiente de Jaccard e distância de Leveshtein - e similaridade de clusterização de redes sociais. No entanto, foi verificado que mesmo após a utilização de várias estratégias, ainda existem nomes de autores que estão ambíguos na base bibliográfica utilizada e isso pode acontecer por alguns motivos, onde podemos citar:

- Nomes de autores muito extensos, com vários nomes intermediários e muitas abreviações;
- Utilização de estratégia errada para desambiguação;
- Utilização de uma estratégia correta, porém em um segundo momento o algoritmo detecta um nome desambiguado como um verdadeiro-negativo e aplica outra técnica de desambiguação, gerando ambiguidade novamente.

Seguindo essa metodologia de desambiguação multi-estratégica, temos o seguinte problema: Qual a melhor estratégia para desambiguação de nomes? Em qual etapa da abordagem multi-estratégica certo nome já está desambiguado na base?

Assim, a proposta para o projeto desta disciplina e resolução do problema proposto é a utilização de um Sistema Multiagentes que consiga definir uma melhor estratégia para desambiguação de nomes utilizando a metodologia multi-estratégica citada anteriormente. Essa escolha da certa estratégia poderá diminuir os casos de nomes falso-positivos e consequentemente aumentar a quantidade de nomes desambiguados, em comparação com a metodologia citada.

A seguir serão definidos os agentes e ambiente que serão implementados no projeto.

## 2 Agentes e PEAS

### 2.1 Agente Mestre - PEAS

- **P** (performance): O agente irá detectar uma nova entrada de autor (documento com informações de publicações e coautorias). Irá então buscar na base se existe ambiguação com relação à esse autor, se existir buscar a melhor estratégia para ser utilizada.

- **E** (environment):

A base bibliográfica que está sendo criada, contendo informações de autores, co-autores, publicações e instituições. O tamanho do ambiente pode variar de nenhum registro até o total de registros de todas as instituições utilizadas (5 instituições).

#### Caracterização:

- **Parcialmente observável:** alguns estados do ambiente não são perceptíveis, tais como novos registros de publicações e autores. Podem existir novas publicações que não foram carregadas no banco de dados até o momento de percepção e ação.
  - **Não determinístico:** as ações dos agentes podem gerar possíveis resultados, onde não foram determinadas probabilidades para cada ação acontecer. O agente pode escolher determinada ação de acordo com o tipo de desambiguação necessária.
  - **Sequencial:** As decisões do agente afetam as decisões futuras, por exemplo, se for escolhida uma estratégia de desambiguação errada, em outro momento pode ser escolhida outra estratégias de desambiguação com base em um nome desambiguado incorretamente e então gerar um efeito cascata.
  - **Estático:** O ambiente descrito não se modifica enquanto o agente está tomando uma decisão. Enquanto o agente não escolhe a estratégia correta de desambiguação não são inseridos novos registros de autores.
  - **Contínuo:** Possui uma quantidade infinita de estados, ou seja, enquanto tiver novas publicações e autores, novos registros são inseridos no banco de dados.
  - **Multiagente:** Possui dois agentes. Um agente mestre que escolhe a melhor estratégia a ser adotada. Outro agente que é requisitado e executa a estratégia solicitada.
- **A** (actuators):
    - Ler (Documento de Autor com publicações e co-autores);
    - Acesso ao banco de dados;
    - Designação estratégia;

- Comunicação com agente executor;
- **S** (sensors):
  - Fila de inserção de autores (Detecta um novo autor para inserir e verificar necessidade de desambiguação);
  - Detecção confirmação de término de procedimento de desambiguação pelo agente executor.

## 2.2 Agente Executor de Estratégia - PEAS

- **P** (performance): O agente receberá informação do agente mestre com o registro de nome de autor que irá ser desambiguado e o tipo de desambiguação que ele aplicará.

Após desambiguação retornará ao agente mestre a informação de término de desambiguação.

- **E** (environment):

A base bibliográfica que está sendo criada, contendo informações de autores, co-autores, publicações e instituições. O tamanho do ambiente pode variar de nenhum registro até o total de registros de todas as instituições utilizadas (5 instituições).

### Caracterização:

- **Parcialmente observável:** alguns estados do ambiente não são perceptíveis, tais como novos registros de publicações e autores. Podem existir novas publicações que não foram carregadas no banco de dados até o momento de percepção e ação.
- **Não determinístico:** as ações dos agentes podem gerar possíveis resultados, onde não foram determinadas probabilidades para cada ação acontecer. O agente pode escolher determinada ação de acordo com o tipo de desambiguação necessária.
- **Sequencial:** As decisões do agente afetam as decisões futuras, por exemplo, se for escolhida uma estratégia de desambiguação errada, em outro momento pode ser escolhida outra estratégias de desambiguação com base em um nome desambiguado incorretamente e então gerar um efeito cascata.
- **Estático:** O ambiente descrito não se modifica enquanto o agente está tomando uma decisão. Enquanto o agente não escolhe a estratégia correta de desambiguação não são inseridos novos registros de autores.
- **Contínuo:** Possui uma quantidade infinita de estados, ou seja, enquanto tiver novas publicações e autores, novos registros são inseridos no banco de dados.

- **Multiagente:** Possui dois agentes. Um agente mestre que escolhe a melhor estratégia a ser adotada. Outro agente que é requisitado e executa a estratégia solicitada.
- **A** (actuators):
  - Acessar banco de dados;
  - Aplicação de metodologia (comparação de strings, remoção de acentos, análise de similaridade de clusterização de redes);
  - Comunicar com agente mestre;
- **S** (sensors):
  - Recepção de nova informação de agente mestre (qual nome de autor será desambiguado, qual metodologia utilizar);
  - Identificador de término de aplicação de metodologia;