



Título: Autoencoder baseado em Transformers para Blocking em Resolução de Entidades.

Data: 14/07/2022

Horário: 10h00

Local: Videoconferência

Resumo:

Com tantas fontes geradoras de dados na atualidade, é comum que várias dessas fontes possuam suas próprias representações de entidades do mundo real. Saber identificar padrões

de diferentes fontes relacionados a uma mesma entidade não é uma tarefa simples, principalmente, quando se trata de dados em larga escala. Para lidar com esse problema, existe um conjunto de técnicas que buscam resolver as entidades nas diferentes fontes e encontrar os seus respectivos padrões. A resolução de entidades possui duas etapas principais: blocking e matching, sendo a primeira responsável por tornar o processo escalável. O presente trabalho propõe um novo modelo baseado em Transformers e autoencoders para geração de blocos durante a resolução de entidades de dados textuais. Mais especificamente, a proposta busca ter melhor efetividade em dados cuja diferença presente nos textos seja semântica e contextual. A avaliação do modelo será feita com duas métricas já aplicadas na literatura: recall e P/E ratio, e com dois grupos de datasets: o grupo (1) contendo datasets bastante utilizados na literatura, porém, com diferenças entre os textos não baseadas em contexto, e o grupo (2) em que a interpretação do contexto é crucial para identificação das duplicatas.

Banca examinadora:

- Prof. Dr. José Antônio Fernandes de Macêdo (MDCC/UFC - Orientador)
- Prof. Dr. Regis Pires Magalhães (UFC - Coorientador)
- Prof. Dr. Ticiania Linhares Coelho da Silva (UFC)
- Prof. Dr. João Paulo Pordeus Gomes (UFC)