



Projeto de Formatura – 2020 – Press Release

PCS - Departamento de Engenharia de Computação e Sistemas Digitais

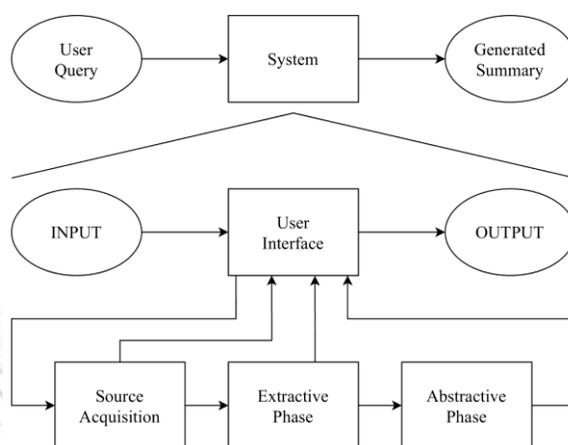
Engenharia Elétrica – Ênfase Computação

Tema:

Arquitetura para um sistema de sumarização abstrativa de múltiplos documentos

Estudantes da Escola Politécnica da Universidade de São Paulo propõem arquitetura para um sistema que gera resumo de múltiplos documentos sobre um determinado tópico automaticamente

Nos dias 5 e 6 de janeiro de 2022, os alunos Lucas Antunes Rodrigues e Matheus Henrique de Sousa Vieira apresentarão o seu trabalho de formatura no curso de Engenharia Elétrica com Ênfase em Computação, desenvolvido sob orientação da Professora Doutora Anna Helena Reali Costa.



O projeto teve como objetivo desenhar e construir um sistema capaz de gerar resumos sobre um determinado tópico automaticamente, com base em uma consulta previamente escolhida pelo usuário e no acesso a múltiplos documentos disponíveis na internet. Assim, os estudantes visaram propor uma arquitetura para este sistema a fim de selecionar, testar e unir tecnologias necessárias, e implementar um protótipo minimamente viável para demonstrar o trabalho.

A Wikipédia é uma enciclopédia livre que hospeda mais de 55 milhões de artigos em centenas de idiomas. No entanto, ainda carece de artigos em diferentes assuntos do conhecimento humano. Uma possível solução para este problema é tomar vantagem da ampla quantidade de informação disponível na internet de forma automática. A sumarização abstrativa de múltiplos documentos relevantes da web pode auxiliar na construção de textos para a enciclopédia livre, bem como para outras atividades textuais do dia a dia.

Após implementar a arquitetura do sistema, a equipe pode avaliar seu desempenho sob diferentes tipos de consultas e cenários de documentos a fim de identificar pontos de sucesso e de melhoria. Unir documentos de fontes distintas (PDFs, páginas web e arquivos texto) provou ser de enorme riqueza na construção de um corpus sobre determinado assunto. Além disso, a extração de informação através de busca semântica demonstrou capacidade considerável de abstração do conteúdo dos textos. Por fim, o sistema desempenhou sua função de forma rápida, geralmente concluindo o processo em menos de dois minutos, algo que um ser humano dificilmente é capaz de realizar caso exerça a mesma função.

Integrantes: Lucas Antunes Rodrigues
Matheus Henrique de Sousa Vieira

Professor(a) Orientador(a): Anna Helena Reali Costa