

# **Apresentação TCC 1**

**Filipe Nunes Soares,<sup>1</sup>**

<sup>1</sup>Instituto de Ciências Exatas e Aplicadas(ICEA) - Universidade Federal de Ouro Preto  
R. 36, 115 - Bairro Loanda - CEP 35931-008 – João Monlevade – MG – Brasil

<sup>2</sup>Departamento de Computação e Sistemas(DECSI)

## **1. Introdução**

A acessibilidade web é fundamental para garantir que todos possam participar plenamente da vida digital, independentemente de suas capacidades físicas, cognitivas ou sensoriais.

Para alcançar a acessibilidade web, os desenvolvedores devem seguir algumas recomendações como as do e-MAG.

## **2. Justificativa**

A ferramenta que iremos utilizar é a ASES.

Ela permite que um usuário insira um link, faça o upload de um arquivo ou código fonte a ser avaliado e identifique os erros de acessibilidade com base nos parâmetros internos estabelecidos pela ferramenta e mostra a porcentagem de conformidade de acordo com o e-MAG.

No entanto, essa ferramenta só pode avaliar a página indexada, o que impede a verificação da experiência real do usuário com deficiência em relação à acessibilidade do site como um todo.

É difícil avaliar todos os sites de um domínio, especialmente quando a instituição não tem controle total sobre todos os subdomínios e links externos associados às páginas registradas, como é o caso da Universidade Federal de Ouro Preto (UFOP).

## **3. Referencial teórico**

A acessibilidade web é um conjunto de práticas de desenvolvimento que têm como objetivo tornar a tecnologia acessível a todos os usuários, incluindo pessoas com deficiência ou estão aquelas incapacitadas um determinado momento. Todos devem ser capazes de utilizar de forma independente a web.

O e-MAG é a recomendação brasileira de acessibilidade na web, criada em 2005, que estabelece normas e critérios para a melhoria da acessibilidade e inclusão de pessoas com deficiência.

Existem três tipos de avaliação de acessibilidade na web: com usuários, com especialistas e com ferramentas de avaliação. O foco deste trabalho é avaliar por meio de ferramentas que percorrem o código para encontrar erros de acessibilidade.

Um web crawler é um tipo de programa que a partir de um link, ele faz o download das páginas, extrai todos os links presentes nessas páginas e, em seguida, continua o processo recursivamente, fazendo o download das páginas identificadas.

## 4. Objetivo

O objetivo deste trabalho consiste em desenvolver um Web Crawler para avaliar a acessibilidade web de todas as paginas do dominio da UFOP.

- Mapear e rastrear todos os links que constam no domínio da UFOP.
- Avaliar a acessibilidade, para cada página encontrada, deverá ser avaliada pela ferramenta ASES.
- Compilar os resultados que foram encontrados pelo ASES.

## 5. Paradigma de construção de conhecimento - Metodologia

Em nosso projeto iremos utilizar o Design Science Research (DSR) que é um paradigma de construção de conhecimento que tem o objetivo de criar coisas uteis com rigor científico, mas iremos utiliza-lo como uma metodologia.

Deixando claro que por conta da complexidade do projeto, possivelmente não terá tempo suficiente para realizar todas avaliações e teste necessários, e será feito posteriormente em trabalhos futuros.

Selecionamos alguns passos para gerar o conhecimento e criar o artefato, sendo elas:

## 6. Resultados parciais

**Para adiantar o desenvolvimento do projeto, já iniciamos a construção do crawler que será responsável por rastrear os link.**

De inicio já conseguimos identificar alguns problema em links para determinar o que será rastreado.

- A minha UFOP foi um deles, pois é necessário autenticação para que o Crawler consiga rastrear as páginas.
- Além disso, como o objetivo é avalias apenas o domínio da UFOP links externos não necessitam ser rastreados.
- PDFs - decidimos não avalia-los pois a ferramenta do ASES, apesar de dar suporte, não é sua principal função.
- A revista coringa foi uma escolha por conta de processamento.
- PNGs, GIFs entre outros não serão rastreados pois as imagens já estão em alguma pagina

### 6.1. Exemplo

**Aqui está um exemplo de como é feito esta coleta e como chegamos aos números que irei apresentar.**

Inicialmente começamos o processo com a pagina inicial da UFOP, então o crawler irá coletar todo o HTML desta pagina e coletará todos os links encontrados, independente se for um PDF, gif, ou pagina invalida. A seguir ele irá passar esse link pelo filtro que criamos e irá salvar em uma lista de processamento.

**Vale ressaltar que links duplicados não entram nessa lista, então caso um link já tenha sido encontrado, ele não será adicionado novamente.**

Como exemplo o link da acessibilidade da pagina inicial.

Apos o processo pegara um desses links validos encontrados, e fara o mesmo processo, até que não tenham mais nenhum link para ser rastreado.

## **6.2. Numeros encontrados**

Entramos em contato com o NTI, e nos foi dito que no domínio da UFOP tem 1450 subdomínios registrados, então para cada uma delas, teremos diversas paginas não sendo possível saber quantas são.

A partir de nossos testes iniciais, que será a amostra considerada por enquanto, deixamos o processo rodando por 16hrs com 10 threads em simultâneo. Com isso foram encontrados + de 460.000 links sendo eles + 12.000 sendo PDFs e uma curiosidade de que + de 1.000 links estavam com erro, sendo eles links que levam para paginas não existentes.

**Uma consideração é de que caso tenhamos 2 links que levam para a mesma pagina, ainda estamos coletando os dois, mas que será tratado no decorrer do projeto. Ex: tem links que por conta do jeito que a página foi feito, tem // ou então so uma / ...**

## **7. Conclusão**

Agora irei abrir o momento para discussão e perguntas, muito obrigado.

## **8. Perguntas de bancas**

### **8.1. Por que o ases?**

O ases é a única ferramenta que avalia de acordo com o e-MAG, e ele são as recomendações para paginas que são do governo, e o site da UFOP é governamental.

### **8.2. Como vou fazer a integração do Ases com o Crawler?**

### **8.3. Como voce vai compilar esses registrados**

### **8.4. Qualquer coisa que eu n souber responder**

A gente dedicou muito tempo agora fazendo o crawler e estudando ele no TCC 1 , e agora no TCC 2 iremos focar em fazer a integração e estudar isso.

### **8.5. Tecnologia**

Python

Biblioteca urllib.request