**Programa DIO**

**Formação HTML Web Developer**

**HTML II – Ampliando o Conhecimento**

Entendendo HTML Semântico

O que significa a palavra semântica?

Disciplina que estuda a significação das palavras e expressões linguísticas bem como das relações de sentido que estas estabelecem entre si.

O HTML semântico – significa que o HTML tem que fazer sentido. Então o HTML está fazendo sentido para aquilo que ele está proposto a fazer.

Então por exemplo, antigamente tinha-se a tag div para separar as coisas na tag body do HTML...e com isso utilizavam-se o atributo class com as denominações variadas, e embora uma pessoa programadora pudesse saber o que significava determinada div, qualquer um poderia colocar qualquer nome da class, ou seja, não havia um padrão.

E o principal problema disso para questões de acessibilidade é que os robôs quando se define classes não estou dando um padrão que os robôs possam ler o HTML.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente Por isso que com a atualização do HTML 4 para o HTML 5 trouxe outras tag mais específicas, principalmente para acessibilidade. Figura abaixo:

Interface gráfica do usuário, Aplicativo, Word

Descrição gerada automaticamente

Assim, se há um padrão nas tags o robô (leitor de tela) conseguirá saber melhor a estrutura do site.

Mudanças do HTML 4 para o HTML 5

Passou de uma visão de formatação para uma visão de estruturação.

Link para estudo :

<https://www.w3.org/TR/html5-diff/>

Acessibilidade

Acessibilidade significa ter um site adaptado para pessoas com deficiência.

A acessibilidade é definida principalmente pelo HTML. Embora o CSS também, mas mais no sentido relacionado a cores.

Portanto, para navegabilidade, geralmente é definido pelo HTML.

Link: https://www.w3c.br/pub/Materiais/PublicacoesW3C/cartilha-w3cbr-acessibilidade-web-fasciculo-I.html

Iniciativa WCAG (Web Content Acessibility Guidelines) da W3C – é um guia para saber como deixar site mais acessível

Link: https://www.w3.org/WAI/WCAG21/quickref/

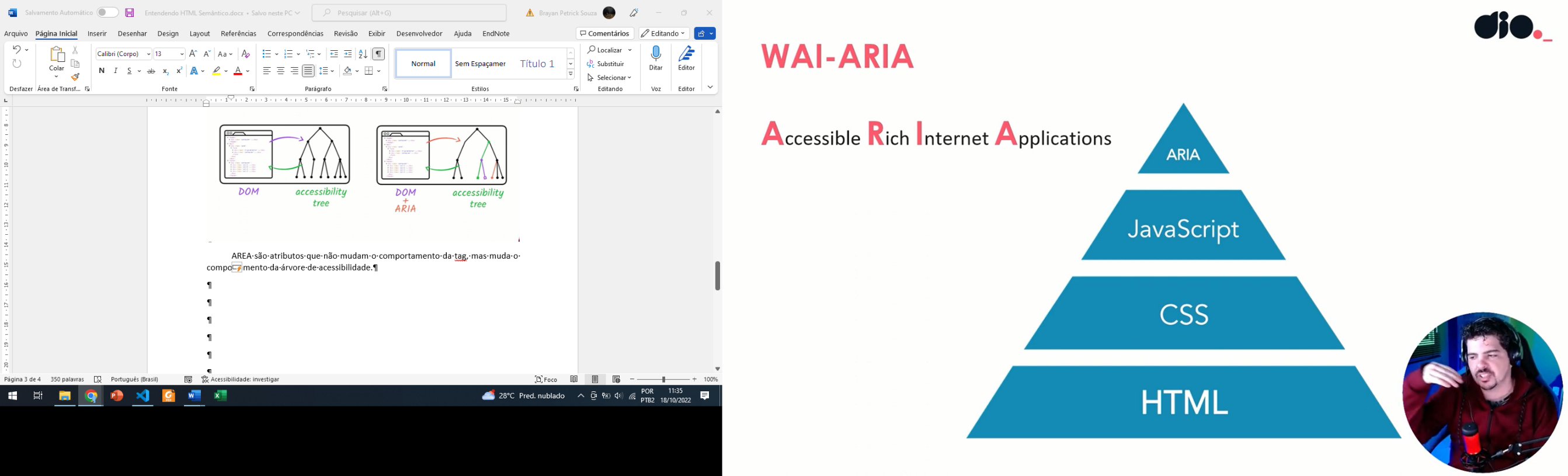
No navegador há leitura do HTML – que se chama de DOM (document object model) que por sua vez irá criar uma árvore de acessibilidade (acessibility tree).

A árvore de acessibilidade representa os níveis de acessibilidade que o usuário com deficiência irá acessar os assuntos que tem na tela dele.

E através o AREA consegue-se manipular a árvore de acessibilidade.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

AREA são atributos que não mudam o comportamento da tag, mas muda o comportamento da árvore de acessibilidade.

Assim, ARIA são atributos que se coloca no HTML de modo que não se altere a execução da tag e sim a árvore de acessibilidade.

Para recolhimento e expansão de um botão de menu:

Aria-expanded, false

Aria-expanded, true

Aria-hidden=”true” – ficará escondido para acessibilidade- ou seja, robô irá ler novamente quando expandir o menu.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Após utiliza-se

Aria-expanded=”false” – pois ao expandir menu ele irá carregar uma página falsa (ou seja, o próprio menu). Assim, a pessoa que não enxergar irá saber que o menu foi aberto, pois, o robô entenderá que o menu foi aberto.

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Por fim, ao clicar e exibir irá dizer que abriu o menu, executando o else, e vice versa clicou esconder vai ler o if. Ou seja, variável só para acessibilidade.

Tela de computador

Descrição gerada automaticamente

Web Scraping

No início os **Crawlers** ou **Rastreadores** também chamado de robô ou indexador, são termos genéricos para qualquer programa usado para descobrir sites automaticamente seguindo links entre as páginas da Web.

Ou seja, é um programa (uma aplicação desenvolvida) que foi programado para entra em um site ler o HTML e extrair as informações e gravam elas em algum local.

Ex. hoje em dia com a linguagem Python, faz-se o seguinte: se cria uma aplicação para que extraia informações mais filtradas de determinados sites e as colocam e uma forma esmiuçada em outra fonte de dados. Seja em Excel, arquivos de texto, bancos de dados relacional e não relacional, etc.

Assim, esses são os Crawlers, e existe um famoso que o **GoogleBot** que faz a indexação das páginas web dentro da busca do Google. Também há a possiblidade pelo Google Search onde se autoriza o Google fazer a varredura do site, e assim começar aparecer no buscador do Google. E é onde entra o SEO (Search Engine Optimization.

Interface gráfica do usuário, Aplicativo, Word

Descrição gerada automaticamente E no buscador do Google aparece assim a página:

Primeiro o Google traz o link do site, trouxe o título da página com link, uma descrição e alguns links complementares embaixo.

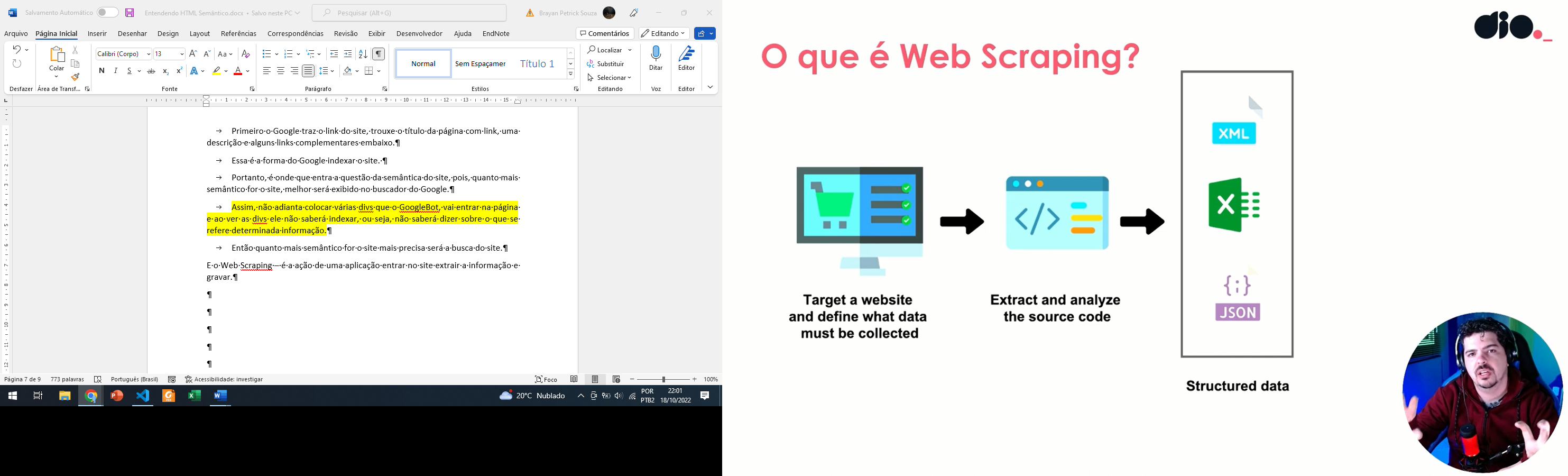
Essa é a forma do Google indexar o site.

Portanto, é onde que entra a questão da semântica do site, pois, quanto mais semântico for o site, melhor será exibido no buscador do Google.

Assim, não adianta colocar várias divs que o GoogleBot, vai entrar na página e ao ver as divs ele não saberá indexar, ou seja, não saberá dizer sobre o que se refere determinada informação.

Então quanto mais semântico for o site mais precisa será a busca do site.

E o Web Scraping – é a ação de uma aplicação entrar no site extrair a informação e gravar.



**Tag <header>, <main> e <footer>**

Tag <header> - cabeçalho

Tag <main> - tem um detalhe, ela não pode ser filha da tag header, footer, article, etc.

* Assim, a tag main é uma tag principal, ou seja, onde contém o assunto principal da página.
  + Assim, mostra-se para os mecanismos de busca que de fato aquele pedaço específico do código é o conteúdo principal da página.

O main só há 1 por página, ou seja, não pode colocar 2 main na página.

Obs. pode-se colocar header dentro de main, articles, section e footer também.

E dentro de article, section pode-se colocar header e footer

Link: tag main

https://developer.mozilla.org/pt-BR/docs/Web/HTML/Element/main

**Tag <aside>, <section> e <nav>**

Dentro da tag main pode-se ter uma lista de seções.

* **Tag section** é como se fosse a div, diferença que a div é genérica, quando se quer ter semântica no google coloca a tag section para enfatizar buscas dentro do e-commerce por exemplo. Como uma lista de coisas para vender, que aparecerá no site.
  + **Link:**

<https://developer.mozilla.org/en-US/docs/Web/HTML/Element/section>

* + Obs. pode-se pegar uma section e abraçar as sections filhas, como se fosse um resultado de busca para determinado filtro.
  + Obs. se for para apenas utilização de CSS, coloca-se div.
  + Assim a regra é utilizar a section quando uma seção deva aparecer logicamente na estrutura do documento.
* **Tag aside** é quando se quer colocar informações que não tenha a ver com o assunto principal da página. Por exemplo, página de vendas de imóveis, onde do lado (aside) ficaria, “fale com o corretor”, “colocar anúncios de corretoras”, “dar espaço no banner do aside para colocar anúncios”, link de blogs, etc.
  + Não se utiliza aside para textos entre parênteses.
  + **Link:**

https://developer.mozilla.org/pt-BR/docs/Web/HTML/Element/aside

* **Tag nav**  - exclusivo para itens de navegação.
  + **Link**:

https://developer.mozilla.org/pt-BR/docs/Web/HTML/Element/nav

* Obs. ao se colocar uma tag nav no footer, para o SEO fica claro que são links de rodapé.
* Obs. ao se colocar tag nav no header (cabeçalho), subentende que é onde ficará navegação principal.
* Geralmente se criar uma navegação no nav, cria-seuma lista não ordenada, embora também possa ser ordenada. Que também reaproveita para estilização.