

COMPILADORES
TRABALHO SOBRE ANÁLISE LÉXICA
Reconhecimento de URLs (Uniform Resource Locators)

Desenvolva uma **gramática regular** (ou algo próximo) que identifique URLs, incluindo: protocolo (opcional), domínio, porta (opcional), caminho (opcional), parâmetros de consulta (opcional) e fragmentos (opcional). Use **expressões regulares** para identificar cada item das URLs. A gramática deve ser submetida ao ANTLR para ser gerado automaticamente os analisadores léxico e sintático.

As URLs servem para identificar recursos na Internet e são compostas por várias partes distintas, cada uma com sua funcionalidade e sintaxe específica.

- **Protocolo:** indicam o método ou o tipo de serviço utilizado para acessar o recurso na web. Vamos considerar somente os protocolos http, https e ftp.
- **Domínio:** Nome do host ou IP que indica o servidor onde o recurso está localizado.
- **Porta** (opcional): Número da porta no servidor para conexão, padrão é 80 para HTTP e 443 para HTTPS.
- **Caminho:** Caminho no servidor onde o recurso específico está localizado.
- **Query** (opcional): Uma string de consulta que contém parâmetros adicionais para o servidor. Geralmente é formatada como chave=valor e separada por &. Eles são importantes em muitas aplicações web para filtrar conteúdo, identificar sessões, passar configurações, entre outros. Alguns exemplos comuns incluem:
 - ?user_id=123: Identifica um usuário específico.
 - ?page=4: Indica uma página específica nos resultados.
 - ?search=keyword: Passa uma palavra-chave de busca.
 - ?sort=asc&field=price: Instruções para ordenação de conteúdo.
- **Fragmento** (opcional): Uma âncora para ir a uma seção específica dentro do recurso.

Exemplos de URLs válidas

Exemplo 1: URL com HTTP

http://www.example.com:80/path/to/file?search=query&sort=ascending#section2

Protocolo: http

Domínio: www.example.com

Porta: 80

Caminho: /path/to/file

Query: search=query&sort=ascending

Fragmento: section2

Exemplo 2: URL com HTTPS

https://secure.example.com/register?user=abc&token=123#signup

Protocolo: https

Domínio: secure.example.com

Caminho: /register

Query: user=abc&token=123

Fragmento: signup

Exemplo 3: URL com FTP

ftp://example.com/downloads/file.zip

Protocolo: ftp

Domínio: example.com

Caminho: /downloads/file.zip

Observações:

- 1) Execute o Antlr sobre a gramática e **gere** os analisadores léxicos e sintáticos;
- 2) **Gere** um programa que instancie os analisadores léxico e sintático, leia o texto da URL via terminal e imprima a árvore de parse reconhecida;
- 3) **Não é objetivo** verificar se a URL está semanticamente correta. Por exemplo, query e fragmento não são comuns em URLs FTP, mas você aceitará, caso seja fornecido. Não esqueça que a análise efetiva será feita nas análises sintática e semântica;
- 4) As três palavras chaves são **http, https e ftp**;
- 5) Considere que as palavras chaves **não devem aparecer** nos identificadores de domínio, caminho, query e fragmento;
- 6) Entregue somente **dois arquivos**: a gramática e o programa em Python.