

Bison

Zadanie polega na stworzeniu **parsera** wąskiego **podzbioru HTML-a** ([HTML 4.01 Specification](#)).

Proszę wziąć pod uwagę następujące znaczniki:

- nagłówki `<h1></h1>`, `<h2></h2>`, `<h3></h3>`
- link ``
- paragraph, bold, break `<p></p>`, ``, `
`

Dla przykładowego poprawnego pliku:

```
<html>
  <head>
    <title>test</title>
  </head>
  <body>
    <h1>Opis cwiczenia</h1>
    <p>Zadanie polega na stworzeniu <b>parsera</b> podzbioru
    <b><a href="http://www.w3.org/TR/html4/">HTML-a</a></b>.</p>

    <h3>Jak wyslac wyniki?</h3>
    <p>Rozwiazanie nalezy przeslac z wykorzystaniem skryptu<br>
    <a href="wyslij-wtB-lab2">wyslij-wtB</a></p>
  </body>
</html>
```

na wyjściu powinien pojawić się komunikat:

```
Poprawny plik
```

W przypadku błędu w użyciu znacznika lub pojawienia się znacznika innego od wymienionych proszę o wypisać komunikat i kontynuować analizę.

Przykładowo dla niepoprawnego wejścia:

```
<html>
  <body>
    Inny opis cwiczenia</h2>
    <p>Zadanie polega na stworzeniu parsera podzbioru
    <b><a href="htt://www.w3.org/TR/html4/">HTML-a</a></b>.</p>
    <p>Rozwiazanie nalezy przeslac z wykorzystaniem skryptu
    <a>wyslij-wtB</a> z konta <i>na studencie</i>.</p>
    <p>Inne materialy
  </body>
</html>
```

na wyjściu powinny pojawić się komunikaty o błędach np. w następującej formie:

```
Bład w linii 3: niesparowany znacznik </h2>
Bład w linii 4: nieprawidłowy format znacznika <a>wyslij-wtB</a>
Bład w linii 5: nieprawidłowy format znacznika
    <a href="htt://www.w3.org/TR/html4/">HTML-a</a>
Bład w linii 5: nieznaný znacznik <i></i>
Bład w linii 6: niesparowany znacznik <p>
```

Zakładamy, że na wejściu parsera nie pojawiają się polskie znaki. Znaczniki `<head></head>` i `<title></title>` są opcjonalne. Proszę sprawdzić poprawność adresów url. Dla ułatwienia może założymy, że url rozpoczyna się od http, https, ftp i

file. W url-path mogą pojawiać się znaki typu [a-zA-Z0-9.:_-] rozdzielone /. Proszę sprawdzić poprawność adresu URL. Dla ułatwienia może rozpoczynać się od http, https i ftp. W url-path mogą pojawiać się znaki typu [a-zA-Z0-9.:_-] rozdzielone /. Nie trzeba obsługiwać zgnieżdżeń znaczników <p></p>. Dla uproszczenia zakładamy, że wymóg domykania tych znaczników (inaczej niż w standardzie HTML 4.01).

W zadaniu należy wykorzystać generator parserów bison ([bison manual](#)).
Inne materiały: "[Yacc: Yet Another Compiler-Compiler](#)", S.C. Johnson