
Pàgina HTML correcta

Z84907_ca

Aquest problema està basat en el problema "Anti Cross-Site Scripting" del D1, a on es definia una pàgina HTML com a una seqüència de tags i paraules. El text original de la definició es troba al final d'aquest enunciat.

Es vol determinar si una seqüència de paraules és una **pàgina HTML correcta**. A més de la correspondència entre els tags, volem ser més fidels a l'estàndar HTML i aplicar dues condicions noves:

- Que hi hagi un sol *tag* `<html>` que engloba tots els altres.
- Que la seqüència no estigui formada per dues o més seqüències HTML correctes concatenades, com ara `<html> A </html> <html> B </html>`

Així doncs, fes la funció:

```
/**
 * @brief Determina si una seqüència de paraules és una pàgina
 *        HTML correcta
 *
 * @param in Un 'stream' d'entrada d'on llegir les paraules
 *          (S'utilitza igual que 'cin')
 *
 * @returns 'true' si la seqüència és una pàgina HTML correcta,
 *         'false' altrament.
 */
bool is_correct_html(istream& in);
```

Avís: és molt important llegir les paraules del paràmetre `in` en comptes de `cin`. Es dona un exemple comentat en el fitxer `main.cc`.

Observació

Els fitxers públics (icona del gatet) contenen:

<code>main.cc</code>	el programa principal, amb la entrada/sortida feta
<code>html_elem.cc</code>	un mòdul d'utilitats (el mateix que al D1)
<code>Makefile</code>	per compilar amb <code>make</code> còmodament
<code>.vscode</code>	carpeta per compilar i debuggar amb VSCode

Cal implementar `is_correct_hml` en un **fitxer** `.cc` **nou**, compilar, i finalment **enviar** **només el fitxer amb la funció**.

Entrada

Vàries seqüències d'HTML, amb tags i text. Cada seqüència ocupa una línia de text diferent. (La lectura línia per línia ja la fa el programa principal.)

Sortida

Per a cada seqüència, la paraula "HTML" si és una pàgina HTML correcta, o bé "text" si és un simple text. (De l'escriptura ja se n'ocupa el programa principal utilitzant el valor retornat per `is_correct_html`)

Exemple d'entrada

```
<html> </html>
<html> a b c </html>
<html> <a> </a> <b> </b> </html>
<html> <a> c <b> </b> </a> d </html>
oh!
<div> x </div>
<html> y </div>
</html> z <html>
<html> </html> 1 <html> 2 </html>
<html> 3 </a> 4 <a> </html>
<div> 5 </html>
```

Exemple de sortida

```
HTML
HTML
HTML
HTML
text
text
text
text
text
text
text
```

Característiques d'una pàgina HTML (del problema "Anti Cross-Site Scripting")

Una seqüència HTML està formada per paraules de dos tipus: paraules que són *tags* d'HTML, i paraules que no ho són.

Els *tags* HTML sempre tenen un nom delimitat per "<" i ">" i tenen dues versions: d'**obrir** (<nom>) i de **tancar** (</nom>). Noteu que els de tancar tenen una barra (/) com a segon caràcter. Es pot pensar que una parella de *tags* d'obrir i tancar són com dos "parèntesis" d'obrir i tancar, però **amb nom**, de tal manera que els podem distingir entre ells. Per exemple, l'equivalent a "<a>" seria "(())", però en el cas dels *tags* podem distingir entre a i b.

Les paraules d'entrada formen una pàgina HTML correcta quan tots els *tags* d'obrir tenen un *tag* de tancar corresponent, tal com passa amb els parèntesis en una expressió ben parentitzada, però aquest cop amb **correspondència de nom**. Per exemple, les seqüències següents, una per línia, són pàgines HTML correctes:

```
<html> <head> </head> <body> </body> </html>
<html> <head> <t> </t> </head> <body> <a> </a> </body> </html>
<html> <p> <span> </span> </p> <ul> <li> </li> </ul> </html>
```

Informació del problema

Autor : Pau Fernández

Generació : 2025-04-02 21:14:58

© Jutge.org, 2006–2025.

<https://jutge.org>