

---

## Corregeix Pàgines HTML

W54048\_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.*

Els navegadors accepten HTML incorrecte i simplement el **corregeixen** per ser més amables amb els usuaris.

Corregir una pàgina HTML vol dir transformar la seqüència de *tags* i paraules rebuda i produir-ne una de nova. Normalment, la majoria d'elements de la seqüència es deixen tal com estan, però eventualment es poden requerir modificacions com ara: ometre certs *tags* (esborrar-los); o insertar *tags* nous.

Així doncs, implementa la funció `fix_html`, que té les modificacions a fer documentades en Doxygen:

```
/**
 * @brief Corregeix una seqüència de paraules que representa
 *        una pàgina HTML.
 *
 * Cal fer 3 correccions:
 * 1. Si un tag de tancar té un nom diferent del tag d'obrir
 *    corresponent, cal substituir-lo pel correcte.
 * 2. Si un tag de tancar no té el seu tag corresponent d'obrir,
 *    cal esborrar-lo (és a dir, no afegir-lo al resultat).
 * 3. Si al final de la seqüència han quedat tags oberts, cal
 *    afegir els tags de tancar corresponents.
 *
 * @param in Un 'stream' d'entrada d'on llegir les paraules
 *          (S'utilitza igual que 'cin')
 *
 * @returns Un 'vector' amb la seqüència HTML corregida
 */
vector<string> fix_html(istream& in);
```

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

### 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 cas, es mostra a la sortida la seqüència corregida que ha produït la funció `fix_html`, en una línia. (De l'escriptura ja se n'ocupa el programa principal.)

## Observació

Els fitxers públics (icona del gatet) contenen:

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

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

### Exemple d'entrada

```
<html> </html>
<body> a b c </header>
<div> <a> </b> <b> </a> </div>
oh!
<div> x </div>
<html> y </div>
</html> z <html>
<html> </html> 1 <html> 2 </html>
<html> 3 <a> 4 </b> <c> </d> </html>
<div> 5 <html> 6
```

### Exemple de sortida

```
<html> </html>
<body> a b c </body>
<div> <a> </a> <b> </b> </div>
oh!
<div> x </div>
<html> y </html>
z <html> </html>
<html> </html> 1 <html> 2 </html>
<html> 3 <a> 4 </a> <c> </c> </html>
<div> 5 <html> 6 </html> </div>
```

## 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><b></b></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 par-entitzada, 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 : 2025-04-03 16:00:10

© Jutge.org, 2006–2025.

<https://jutge.org>