# **Detect HTML Tags**

# HackerRank

In this challenge, we're using regular expressions to detect the various *tags* used in an HTML document.

- Tags come in pairs. Some tag name, t, will have an opening tag, <t>, followed by some intermediate text, followed by a closing tag, </t>
   The forward slash in a closing tag will always come *before* the tag name.
- The exception to this is *self-closing* tags, which consist of a *single tag* (not a pair) with a forward slash *after* the tag name:

Here are a few examples of tags:

- The **p** tag is for paragraphs: **This is a paragraph**
- There may be 1 or more spaces before or after a tag name:
   This is also a paragraph
- A void or empty tag involves an opening and closing tag with no intermediate characters:

Some tags can also have *attributes*, such as the **a** tag, which is used to add a hyperlink to another document: **<a href="http://www.google.com">Google</a>** 

In the above case, **a** is the tag name and **href** is an attribute having the value **http://www.google.com**.

#### Task

Given N lines of HTML, find the tag names (ignore any attributes) and print them as a single line of lexicographically ordered semicolon-separated values (e.g.: tag1;tag2;tag3).

### **Input Format**

The first line contains an integer, N, the number of HTML fragments. Each of the N subsequent lines contains a fragment of an HTML document.

### Constraints

- $1 \le N \le 100$
- Each fragment contains < 10000 ASCII characters.
- The fragments are chosen from Wikipedia, so analyzing and observing their markup structure may help.
- Leading and trailing spaces/indentation have been trimmed from the HTML fragments.

### **Output Format**

Print a single line containing *all* of the unique tag names found in the input. Your output tags should be semicolon-separated and ordered lexicographically (i.e.: alphabetically). Do not print the same tag name more than once.

### Sample Input

```
2
<a href="http://www.quackit.com/html/tutorial/html_links.cfm">Example Link</a>
<div class="more-info"><a href="http://www.quackit.com/html/examples/html_links_examples.cfm">More Link
Examples...</a></div>
```

## Sample Output

a;div;p

#### Explanation

The first line contains 2 tag names:  $\{\mathrm{p},\mathrm{a}\}.$ 

The second line contains 2 tag names:  $\{div,a\}.$ 

Our set of unique tag names is  $\{p,a,div\}.$ 

When we order these alphabetically and print them as semicolon-separated values, we get "a;div;p".