Build a Stack Exchange Scraper

Stack Exchange is an information power-house, which contains libraries of crowdsourced problems (with answers) across a large number of topics which are as diverse as electronics, cooking , programming, etc.

We are greatly interested in crawling and scraping as many questions, as we can, from stack-exchange. This is an example of a question library page from stack-exchange.

Your task will be, to scrape the questions from each library page, in the order in which they are listed. You will be provided with the markup of question listing pages, from which you need to detect:
(1) Identifier (2) Question text (which is on the Hyperlink to the question) (3) How long ago the question was asked.

The Markup in the Test Cases will be similar to the sample fragment shown below. **Please note, that since this markup is real markup from the website, it is likely to contain some stray control and escape characters, unexpected whitespaces and newlines.**

Sample Markup Fragment

```
      <div class="question-summary" id="question-summary-80407">
      <div class="statscontainer">
        <div class="statsarrow"></div>
        <div class="stats">
          <div class="vote">
            <div class="votes">
              <span class="vote-count-post "><strong>2</strong></span>
              <div class="viewcount">votes</div>
            </div>
          </div>
          <div class="status answered">
            <strong>1</strong>answer
          </div>
        </div>



   <div class="views " title="60 views">
                60 views
   </div>
      </div>
      <div class="summary">
        <h3><a href="/questions/80407/about-power-supply-of-opertional-amplifier" class="question-hyperlink">about power
supply of opertional amplifier</a></h3>
        <div class="excerpt">
          I am constructing an operational amplifier as shown in the following figure. I use a batter as supplier for the OP Amp and
set it up as a non-inverting amp circuit. I saw that the output was clipped ...
        </div>

        <div class="tags t-op-amp">
          <a href="/questions/tagged/op-amp" class="post-tag" title="show questions tagged 'op-amp'" rel="tag">op-amp</a>

        </div>
        <div class="started fr">


      <div class="user-info ">
        <div class="user-action-time">


              asked <span title="2013-08-27 21:49:14Z" class="relativetime">11 hours ago</span>
        </div>
        <div class="user-gravatar32">
          <a href="/users/17060/user1285419"><div class=""><img
src="https://www.gravatar.com/avatar/08ee68b20a4eceff26f7eee99b708c08?s=32&d=identicon&r=PG" alt="" width="32"
height="32"></div></a>
        </div>
        <div class="user-details">
          <a href="/users/17060/user1285419">user1285419</a><br>
          <span class="reputation-score" title="reputation score" dir="ltr">165</span><span title="5 bronze badges"><span
```

```
class="badge3"></span><span class="badgecount">5</span></span>
        </div>
      </div>

        </div>
      </div>
   </div>

   <div class="question-summary" id="question-summary-80405">
      <div class="statscontainer">
         <div class="statsarrow"></div>
         <div class="stats">
            <div class="vote">
               <div class="votes">
                  <span class="vote-count-post "><strong>4</strong></span>
                  <div class="viewcount">votes</div>
               </div>
            </div>
            <div class="status answered-accepted">
               <strong>2</strong>answers
            </div>
         </div>



   <div class="views " title="64 views">
                  64 views
   </div>
      </div>
      <div class="summary">
         <h3><a href="/questions/80405/5v-regulator-power-dissipation" class="question-hyperlink">5V Regulator Power
Dissipation</a></h3>
         <div class="excerpt">
            I am using a 5V regulator (LP2950) from ON Semiconductor. I am using this for USB power and I'm feeding in 9V from an
adapter. USB requires maximum of 500mA right? So the maximum power dissipation in ...
         </div>

         <div class="tags t-voltage-regulator t-surface-mount t-heatsink t-5v t-power-dissipation">
            <a href="/questions/tagged/voltage-regulator" class="post-tag" title="show questions tagged 'voltage-regulator'"
rel="tag">voltage-regulator</a> <a href="/questions/tagged/surface-mount" class="post-tag" title="show questions tagged
'surface-mount'" rel="tag">surface-mount</a> <a href="/questions/tagged/heatsink" class="post-tag" title="show questions
tagged 'heatsink'" rel="tag">heatsink</a> <a href="/questions/tagged/5v" class="post-tag" title="show questions tagged '5v'"
rel="tag">5v</a> <a href="/questions/tagged/power-dissipation" class="post-tag" title="show questions tagged 'power-
dissipation'" rel="tag">power-dissipation</a>

         </div>
         <div class="started fr">


      <div class="user-info ">
         <div class="user-action-time">


                  asked <span title="2013-08-27 21:39:31Z" class="relativetime">11 hours ago</span>
         </div>
         <div class="user-gravatar32">
            <a href="/users/10082/david-norman"><div class=""><img
src="https://www.gravatar.com/avatar/8b073417e471077280b3fc5ff2eaf1f7?s=32&d=identicon&r=PG" alt="" width="32"
height="32"></div></a>
         </div>
         <div class="user-details">
            <a href="/users/10082/david-norman">David Norman</a><br>
            <span class="reputation-score" title="reputation score" dir="ltr">322</span><span title="3 silver badges"><span
class="badge2"></span><span class="badgecount">3</span></span><span title="10 bronze badges"><span
class="badge3"></span><span class="badgecount">10</span></span>
         </div>
      </div>

         </div>
      </div>
   </div>
```

Output Format

The output file should contain N lines, where N is the number of questions you have identified in the provided fragment.Each line contains the identifier, question text and (relative) time when the question was asked (with no leading or trailing spaces surrounding each section); separated by semi-colons. The information about the questions in the output file should match with the ordering in the original markup.

Sample Output

```
80407;about power supply of operational amplifier;11 hours ago
80405;5V Regulator Power Dissipation;11 hours ago
```

Explanation

The given markup fragment points to two questions on electronics.stackexchange.com (at the time the markup was noted).

The first question has ID 80407, it is "about power supply of operational amplifier" and it was asked "11 hours ago" (relative to the time when this markup was noted). Search for these values in the given markup fragment to gain a better understanding of where we identified these values from. The second question has ID 80405, it is about "5V Regulator Power Dissipation", and it was asked "11 hours ago" (relative to the time when this markup was noted).

A Note Regarding the Test Cases

The markup in the test cases will resemble the markup fragment provided above, however, each markup fragment might contain a larger number of questions embedded in it. A markup fragment will have no more than 100 questions embedded in it.