

Лабораторная работа N7. Загрузка и анализ содержимого веб-страниц на PHP

Цель работы: Создать PHP-программу, которая загружала бы веб-страницу с сайта gismeteo.ua, анализировала бы ее и выводила информацию о продолжительности светлого времени суток, а также дневной температуре на свою веб-страницу. Добавить эту программу на свой сайт и поставить на нее ссылку с главной страницы сайта. Оценка за выполненную работу может быть повышена при условии предоставления короткого и красивого исходного текста программы.

Порядок выполнения работы:

1. Сайт gismeteo.ua (рисунок 1) предоставляет пользователям прогнозы погоды.

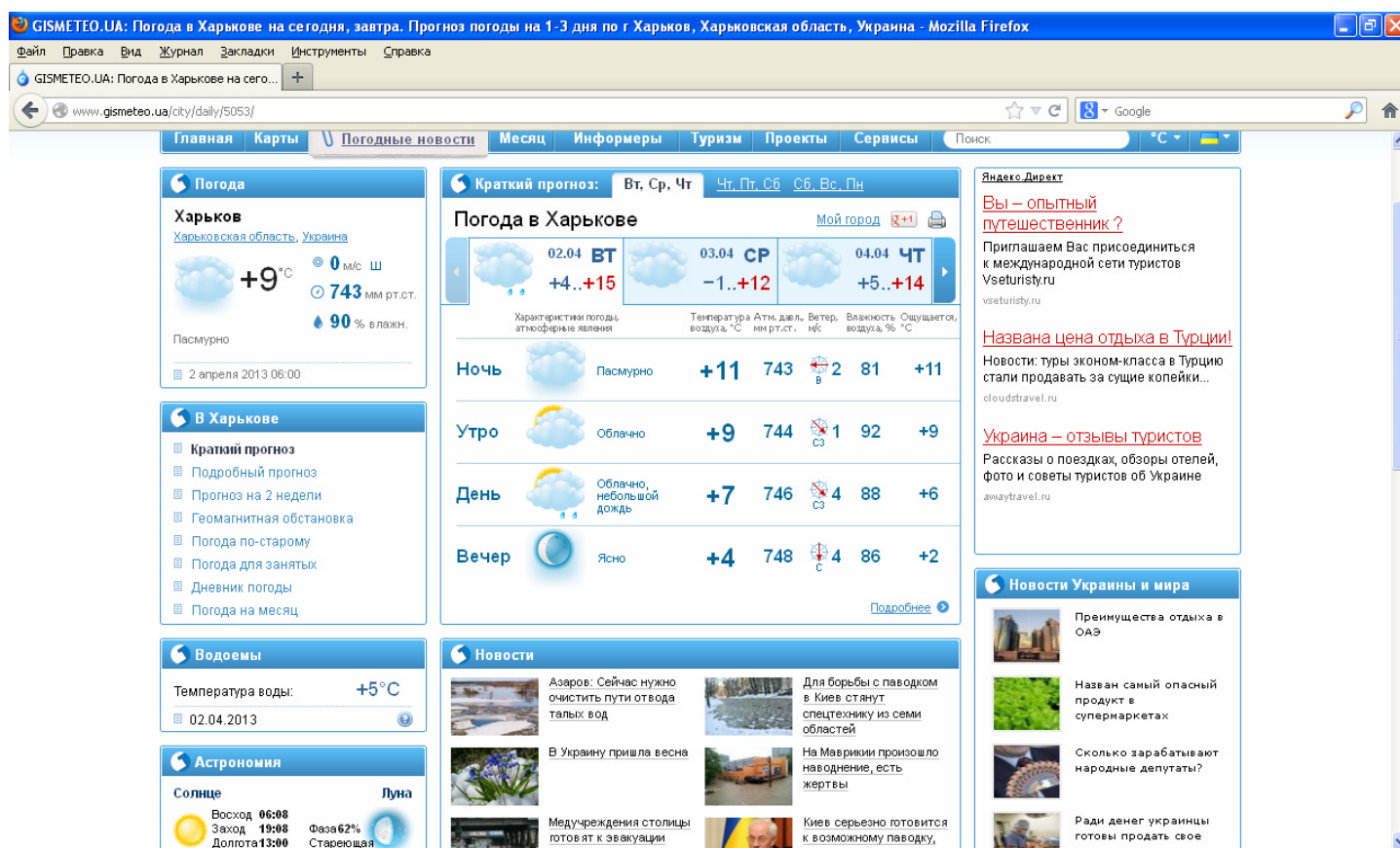


Рисунок 1- Сайт gismeteo.ua

Ваша программа должна загрузить прогноз погоды для города Харькова (ссылка <http://www.gismeteo.ua/city/hourly/5053/>) и вывести на свою веб-страницу следующую информацию: название города (взять из веб-страницы, полученной от gismeteo.ua), текущую дату (тоже взять с сайта), время восхода и захода солнца, длительность светлого времени суток, а также прогноз температуры на 0, 3, 6, 9, 12, 15, 18 и 21 час (см. рисунок 2).

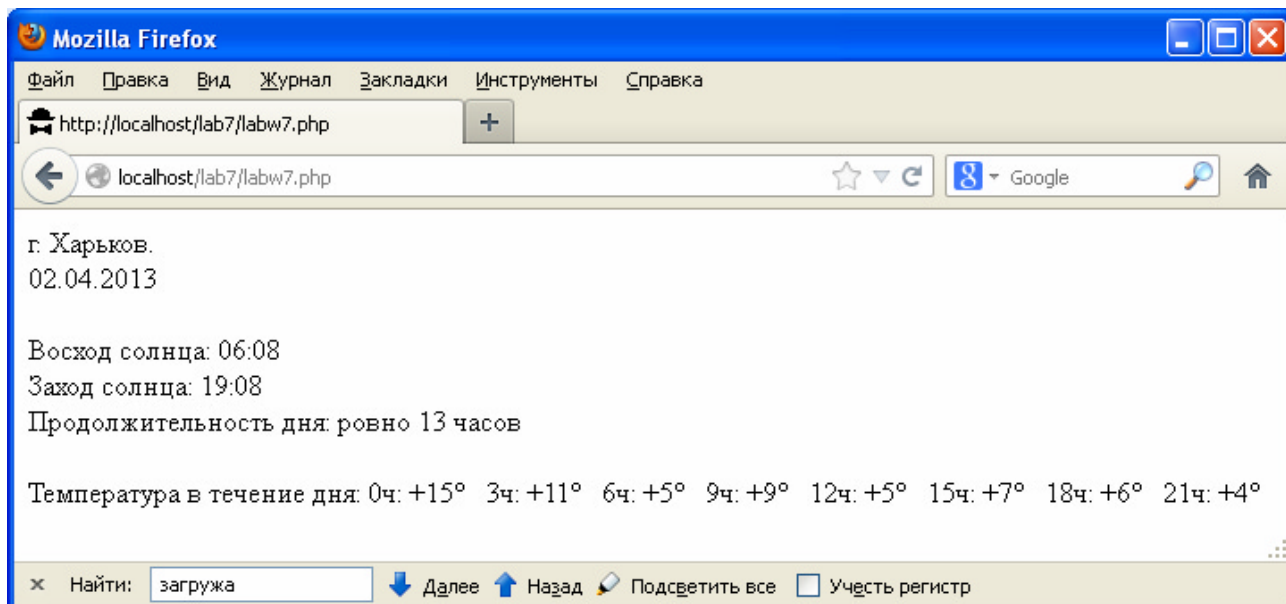


Рисунок 2 - Веб-страница, которая загружает и анализирует информацию с gismeteo.ua

Обратите внимание на то, чтобы ваша программа выводила слова “час”, “часа”, “часов” в нужном падеже в зависимости от конкретной продолжительности дня. Например, “1 час” или “3 часа” или “8 часов”.

Если, как в данном примере, продолжительность дня составляет целое количество часов, то программа должна выводить надпись: “ровно X часов”, где X - количество часов. Если не целое, то выводить надпись: “X часов Y мин”, где Y - количество минут.

2. Замените в своей программе ссылку “<http://www.gismeteo.ua/city/hourly/5053/>” на ссылку “<http://www.gismeteo.ua/city/hourly/4944/>” (прогноз погоды для города Киев). Проверьте, корректно ли работает ваша программа в этом случае (см. рисунок 3).

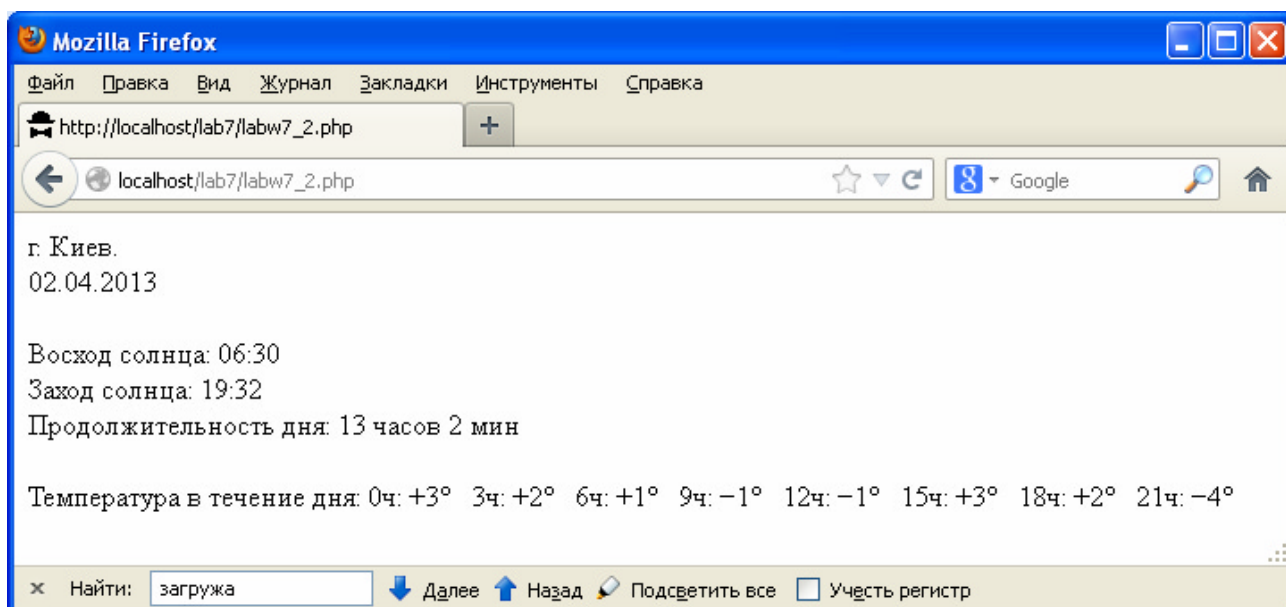


Рисунок 3 - Эта же страница, но обрабатывающая прогноз погоды для г. Киев

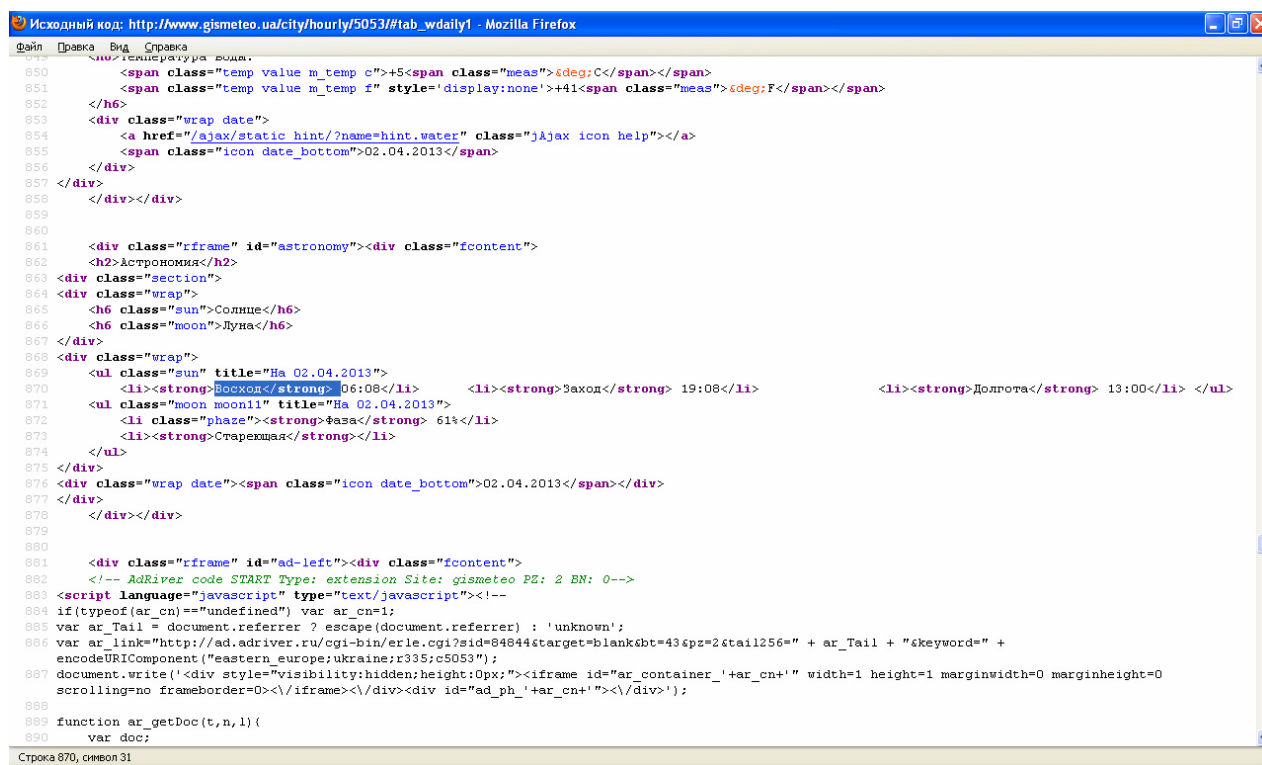
Методические указания:

Для запроса к сайту используйте функции PHP `curl_setopt()`, `curl_exec()` и `curl_close()`. Например, вот так:

```
$curl=curl_init();  
curl_setopt($curl, CURLOPT_URL,  
'http://www.gismeteo.ua/city/hourly/5053/');  
curl_setopt($curl, CURLOPT_RETURNTRANSFER,true);  
$out = curl_exec($curl);
```

Для конвертации полученного содержимого веб-страницы из кодировки UTF-8 в кодировку Windows-1251 используйте функцию `mb_convert_encoding()`.

Выведите на экран исходный код страницы сайта <http://www.gismeteo.ua/city/hourly/5053/> и найдите необходимые ключевые слова (рисунок 4).



```
Исходный код: http://www.gismeteo.ua/city/hourly/5053/#tab_wdaily1 - Mozilla Firefox  
850 <span class="temp value m_temp c">+5<span class="meas"><deg>C</span></span>  
851 <span class="temp value m_temp f" style="display:none">+41<span class="meas"><deg>F</span></span>  
852 </h6>  
853 <div class="wrap date">  
854 <a href="/ajax/static_hint/?name=hint.water" class="ajax icon help"></a>  
855 <span class="icon date_bottom">02.04.2013</span>  
856 </div>  
857 </div>  
858 </div></div>  
859  
860  
861 <div class="iframe" id="astronomy"><div class="fcontent">  
862 <h2>Астрономия</h2>  
863 <div class="section">  
864 <div class="wrap">  
865 <h6 class="sun">Солнце</h6>  
866 <h6 class="moon">Луна</h6>  
867 </div>  
868 <div class="wrap">  
869 <ul class="sun" title="На 02.04.2013">  
870 <li><strong>Восход</strong> 06:08</li> <li><strong>Заход</strong> 19:08</li> <li><strong>Долгота</strong> 13:00</li> </ul>  
871 <ul class="moon moon1" title="На 02.04.2013">  
872 <li class="phase"><strong>фаза</strong> 61%</li>  
873 <li><strong>Старехшая</strong></li>  
874 </ul>  
875 </div>  
876 <div class="wrap date"><span class="icon date_bottom">02.04.2013</span></div>  
877 </div>  
878 </div></div>  
879  
880  
881 <div class="iframe" id="ad-left"><div class="fcontent">  
882 <!-- AdRiver code START Type: extension Site: gismeteo PE: 2 BN: 0-->  
883 <script language="javascript" type="text/javascript"><!--  
884 if(typeof(ar_cn)=="undefined") var ar_cn=1;  
885 var ar_Tail = document.referrer ? escape(document.referrer) : 'unknown';  
886 var ar_Link="http://ad.adriver.ru/cgi-bin/erle.cgi?sid=84844&target=blank&bt=43&pz=2&tail256="+ ar_Tail + "&keyword=" +  
887 encodeURIComponent("eastern_europe;ukraine;r335;c5053");  
888 document.write('<div style="visibility:hidden;height:0px;"><iframe id="ar_container_'+ar_cn+'" width=1 height=1 marginwidth=0 marginheight=0  
889 scrolling=no frameborder=0></iframe></div><div id="ad_ph_'+ar_cn+'"></div>');  
890 function ar_getDoc(t,n,l){  
891 var doc;
```

Рисунок 4 - Пример исходного текста веб-страницы, получаемой от сайта gismeteo.ua

Например, время восхода солнца находится в тексте страницы между фрагментами текста “Восход “ и “”.

Используя регулярные выражения (или написав свою функцию для поиска фрагмента текста) извлеките из страницы всю необходимую Вам информацию.

Например, текст для извлечения из страницы информации о восходе солнца может выглядеть примерно так:

```
$c = preg_match("/Восход<\/strong> .*<li>\/sU", $out,$a);  
$s=$a[0];  
$l=strlen($s);  
$u=substr($s,16,$l-20);  
echo $u;
```

Здесь:

\$out - строка, в которой у нас содержится текст веб-страницы, полученной с gismeteo.ua.

Восход<\/strong> .*\/sU" - регулярное выражение для поиска первого попавшегося фрагмента, начинающегося с “Восход “ и заканчивающегося “” (символы “\” используются здесь для экранирования служебных символов типа “/”, “<”, “>” и др.).

\$s - найденный фрагмент;

\$l - длина строки \$s;

\$u - найденный фрагмент без “Восход “ (16 букв) вначале и “” (4 буквы) вконец (всего - \$l - 20 букв).

По подобию извлеките из веб-страницы всю остальную необходимую информацию (время захода, название города, дату, прогнозы температуры).

Для вывода на экран граудсов используйте спецсимвол HTML “°”.