

Memperoleh Data

PHP memiliki beberapa fitur untuk memudahkan memperoleh data dari sumber eksternal. Dengan `fopen()`, Anda bisa membuka file lokal maupun remote melalui protokol FTP atau HTTP. Dengan berbagai API database, Anda juga bisa meng-query database MySQL, PostgreSQL, Oracle, Interbase, dan sebagainya. Dengan `popen()`, Anda bisa membuka proses atau perintah shell dan mengambil output-nya. Resep-resep pada bab ini berhubungan dengan pengambilan data atau informasi yang sering dibutuhkan sebuah situs dari berbagai sumber.

Resep 2-1: Mengambil Kurs	8
Resep 2-2: Mengambil Headline Berita	12
Resep 2-3: Mengambil Baris Acak dari File	23
Resep 2-4: Mengambil Data dari File CSV	26

Resep 2-1: Mengambil Kurs

Kurs sepertinya sudah menjadi sesuatu yang “wajib hukumnya” bagi banyak situs. Bukan hanya situs e-commerce dan bank saja yang menampilkannya, tapi juga portal berita, homepage ISP, bahkan homepage pribadi pun banyak yang ikut-ikutan memasangnya.

Untuk memperoleh info kurs yang up-to-date, ada beberapa sumber dan versi yang bisa digunakan. Pada resep ini dipakai Yahoo! Finance, <http://finance.yahoo.com>, dan BCA, <http://www.klikbca.com>. Untuk Yahoo! Finance, sintaks URL-nya adalah:

```
http://finance.yahoo.com/m5?a=1&s=<KURS1>&t=<KURS2>
```

di mana <KURS1> adalah simbol mata uang yang dikonversi dan <KURS2> adalah simbol mata uang tujuan konversi. Berikut ini skrip PHP untuk memperoleh kurs. Skrip ini dijalankan dari command line, bukan dari browser.

```
1|#!/usr/bin/php -q
2|<?
3|
4|//
5|// kurs-1.php
6|//
7|
8|if ($argc != 3) {
9|    echo "Cara pakai: kurs-1.php <KURS1> <KURS2>\n";
10|    echo "Contoh: konversi dolar ke rupiah = kurs-1.php USD
    IDR\n";
11|    exit;
12|}
13|
14|$kurs1 = $argv[1]; $kurs2 = $argv[2];
15|
16|if (($fp=@fopen("http://finance.yahoo.com/
    m5?a=1&s=$kurs1&t=$kurs2","r"))
17|    === false)
18|    die("ERROR: Gagal membuka URL!\n");
19|
20|$content = "";
21|while ($chunk = fread($fp, 8192)) $content .= $chunk;
22|
```

```
23|if (!$content) die("ERROR: dokumen URL kosong!\n");
24|
25|if (!preg_match('#<th nowrap>(\d+:\d+[ap]m)</
    th><td>(\d+\.\d+)</td>#',
26|    $content, $m))
27|    die("ERROR: dokumen URL tidak dikenali!\n");
28|
29|$rate = $m[2]; $time = $m[1];
30|
31|echo "1 $kurs1 = $rate $kurs2 ($time)\n";
32|
33|?>
```

Contoh output:

```
$ ./kurs-1.php USD IDR
1 USD = 8500.0000 IDR (10:12pm)
```

Skrip akan menampilkan hasil kurs ke console. Selain mencetaknya ke layar, Anda tentu bisa juga menyimpan hasil kurs ini ke database atau file teks. Sehingga jika digunakan di situs dan pengunjung me-refresh halaman, Anda tak perlu terus-menerus menghubungi server Yahoo!.

Resep kedua di bawah, mengambil kurs dari situs KlikBCA dan menyimpan hasilnya ke file. Di situs KlikBCA terdapat dua halaman yang menyediakan kurs. Pertama di halaman depan (kurs USD, SGD, dan HKD). Kedua, di halaman http://www.klikbca.com/website/indo/price_rate/exchange_rate.html yang berisi daftar lengkapnya. Kita akan mengambil USD, SGD, dan HKD, jadi pilih halaman depan.

```
1|#!/usr/bin/php -q
2|<?
3|
4|//
5|// kurs-bca.php
6|//
7|
8|// tempat menyimpan hasil kurs
9|$SAVE_FILE = "./kurs-bca.txt";
10|
11|if (($fp=@fopen("http://www.klikbca.com/","r")) === false)
12|    die("ERROR: Gagal membuka URL!\n");
```

```

14|$content = "";
15|while ($chunk = fread($fp, 8192)) $content .= $chunk;
16|
17|if (!$content) die("ERROR: dokumen URL kosong!\n");
18|
19|if (!preg_match('#rate.html".+USD.+\\n.+>([\\d,]+\\.\\d+)<#',
    $content, $m))
20|    die("ERROR: Gagal mengekstrak USD dari dokumen!\n");
21|$USD = preg_replace("/,/","",$m[1]); // hilangkan koma
22|
23|if (!preg_match('#rate.html".+SGD.+\\n.+>([\\d,]+\\.\\d+)<#',
    $content, $m))
24|    die("ERROR: Gagal mengekstrak SGD dari dokumen!\n");
25|$SGD = preg_replace("/,/","",$m[1]); // hilangkan koma
26|
27|if (!preg_match('#rate.html".+HKD.+\\n.+>([\\d,]+\\.\\d+)<#',
    $content, $m))
28|    die("ERROR: Gagal mengekstrak HKD dari dokumen!\n");
29|$HKD = preg_replace("/,/","",$m[1]); // hilangkan koma
30|
31|if (($fp = fopen($SAVE_FILE, "w+")) === false)
32|    die("ERROR: Gagal membuka save_file!\n");
33|
34|fputs($fp, "USD=$USD\\nSGD=$SGD\\nHKD=$HKD\\n");
35|fclose($fp);
36|
37|?>

```

Jika berhasil mengambil kurs, skrip tidak akan mengeluarkan output apa-apa ke console melainkan menulis file yang isinya:

USD=8510.00
SGD=4973.20
HKD=1097.05

Karena tidak ada output yang dikeluarkan kecuali pesan error jika gagal, maka skrip ini cocok dijalankan berkala di cron. Misalnya mengambil kurs setiap jam sekali pada menit ke-0:

```
0 * * * * /home/steven/scripts/kurs-bca.php
```

Pembahasan

Kedua resep kurs di atas menggunakan regex dan bukan mem-parse dokumen HTML menggunakan parser, karena:

- 1) Menggunakan regex lebih sederhana dan cepat, apalagi jika data yang ingin diambil hanya sedikit. Pada kasus ini hanya satu atau beberapa buah angka;
- 2) Dokumen HTML situs pada dasarnya “acak-acakan” atau tidak berstruktur, dan kalau situs berubah layout, kita tetap saja harus mengubah pola regex atau parser. Mengubah-ubah pola regex lebih simpel daripada harus mengubah kode parsing.

Untuk menampilkan data file teks kurs yang telah kita ambil dan simpan, bisa digunakan kode seperti berikut:

```

1|<?
|
3|//
4|//  tampilkan-kurs-bca.php
5|//
|
7|$SAVE_FILE = "./kurs-bca.txt";
|
9|echo "<table bgcolor=#999999 cellspacing=0 cellpadding=0
border=0>
10|    <tr><td><table cellspacing=1 cellpadding=5 border=0>;
|
12|if (!file_exists($SAVE_FILE)) {
13|    echo "<tr><td colspan=2 bgcolor=#e0e0e0><font
color=#cc3333><b>
14|        Data kurs tidak ditemukan. </b></font></td></tr>";
|
16|} elseif (($fp = fopen($SAVE_FILE, "r")) === false) {
17|    echo "<tr><td colspan=2 bgcolor=#e0e0e0><font
color=#cc3333>
18|        Data kurs tidak ditemukan. </font></td></tr>";
|
20|} else {
|
22|    echo "<tr><td colspan=2 bgcolor=#333333><font
color=#cccccc><b>
23|        K U R S &nbsp;   </b></font>
24|        <font color=#999999><small>Last updated at
25|        ",date("H:i", filemtime($SAVE_FILE)),"</small></font>

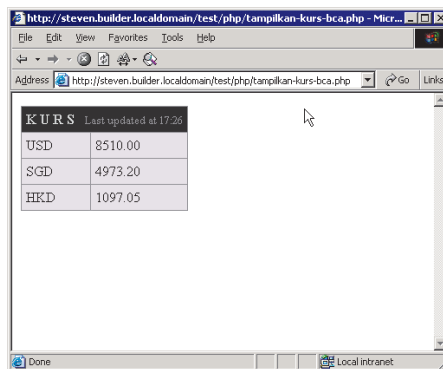
```

```

26|         </td></tr>";
|
28| while ($line = fgets($fp)) {
29|     preg_match("/(\\w+)=\\d+\\.\\d+/", $line, $m);
30|     echo "<tr bgcolor=#e0e0e0><td>$m[1]</td><td>$m[2]</td></tr>";
|
31| }
|
33|}
|
35|echo "</table></td></tr>";
36|     </table>";
|
38|?>

```

Hasilnya bisa dilihat di Gambar 2-1. Skrip ini bisa Anda include, misalnya di halaman-halaman lain, karena berupa kotak yang bisa diselipkan dalam layout.



KURS <small>Last updated at 17:28</small>	
USD	8510.00
SGD	4973.20
HKD	1097.05

Gambar 2-1. Tampilan Kurs.

Resep 2-2: Mengambil Headline Berita

Sebagian situs berita, seperti slashdot.org atau advogato.org, biasanya tidak keberatan jika headline beritanya diambil dan ditampilkan di halaman kita. Malah situs-situs seperti ini membantu menyediakan daftar headline. Daftar headline beritanya disimpan dalam sebuah file XML yang disebut RSS yang sering juga dijuluki “RSS news feed”.

Karena formatnya XML dan standar, maka baik menggunakan regex maupun parser XML, kita dengan mudah dapat memperoleh headline berita. Di PEAR misalnya, tersedia paket XML_RSS untuk memparse dokumen RSS.

Berikut contoh file RSS:

```

<?xml version="1.0" encoding="ISO-8859-1"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns="http://my.netscape.com/rdf/simple/0.9/">

<channel>
  <title>Ruby Garden Wiki</title>
  <link>http://www.rubygarden.org/ruby</link>
  <description>Recent Changes to the Ruby Garden</description>
</channel>

<image>
  <title>Ruby Garden Wiki</title>
  <link>http://www.rubygarden.org/ruby</link>
  <url>http://www.rubygarden.org/ruby.jpg</url>
</image>

  <item>
    <title>WindowsInstaller</title>
    <link>http://www.rubygarden.org/ruby?WindowsInstaller</link>
  </item>

  <item>
    <title>ChrisMorris</title>
    <link>http://www.rubygarden.org/ruby?ChrisMorris</link>
  </item>

  <!-- more items... -->
</rdf:RDF>

```

Sebagian situs lain, seperti detik.com dan kompas.co.id, cenderung tidak terlalu antusias jika headline beritanya ditampilkan di halaman situs kita. Bahkan mereka cenderung protektif untuk mencegah agar isi beritanya tidak bisa diambil. Antara lain dengan trik Javascript. Hal ini dilakukan karena situs-situs seperti ini amat

mengandalkan orang mengunjungi halaman situsnya langsung agar bisa melihat banner. Untuk situs-situs seperti ini, kita perlu mengandalkan regex atau parser khusus. Belum lagi jika berubah layout, kita harus ikut mengubah pola regex atau parser.

Resep di bawah memperlihatkan contoh mengambil headline berita dari Slashdot menggunakan XML_RSS. Situs-situs lain yang menyediakan RSS akan dapat Anda ambil headline-nya dengan cara serupa.



Gambar 2-2. Tampilan Headline.

```

1|<?
2|
3|//
4|// headline-slashdot.php
5|//
6|
7|include 'XML/RSS.php';
8|
9|$rss =& new XML_RSS("http://slashdot.org/slashdot.rdf");
10|$rss->parse();
11|
12|echo "<h3>Headline <a href=\"http://slashdot.org\">Slashdot</a></h3>";
13|echo "<ul>\n";
14|
15|foreach ($rss->getItems() as $item) {
16|    echo "<li><a href=$item[link]>$item[title]</a></li>";
17|}
18|
19|echo "</ul>\n";
20|
21|?>

```

Hasilnya seperti di Gambar 2-2.

Resep di bawah untuk mengambil headline detik.com menggunakan regex:

```

1|<?
2|
3|//
4|// headline-detik.php
5|//
6|
7|if (($fp = fopen("http://www.detik.com/index.htm","r"))===false)
8|    die("Gagal mengambil URL!");
9|
10|$content = "";
11|while ($chunk = fgets($fp)) $content .= $chunk;
12|
13|if (!$content)
14|    die("Dokumen URL kosong!");
15|
16|if (!preg_match_all('#<span class="tanggal(?:High)?">([>]+)</span><BR>' .
17|    '\s*<A href="([>]+)"[>]+><span class="(?:judulHigh|strJudul)">' .
18|    '(.+?)</span></A><BR>\s*<span class="summary(?:High)?">' .
19|    '(.+?)</span>#si', $content, $mm, PREG_SET_ORDER))
20|    die("Gagal mengekstrak headline dari dokumen URL!");
21|
22|$items = array();
23|foreach ($mm as $m) {
24|    $items[] = array(
25|        'date' => $m[1],
26|        'link' => $m[2],
27|        'title' => $m[3],
28|        'summary' => $m[4],
29|    );
30|}
31|
32|echo "<h3>Headline <a href=\"http://www.detik.com\">".
33|    "detikcom</a></h3>";
34|echo "<ul>\n";
35|
36|foreach ($items as $item) {
37|    echo "<li><a href=$item[link]>$item[title]</a></li>";

```

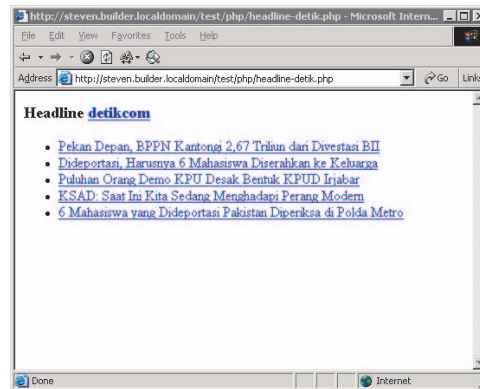
```

38| }
|
40| echo "</ul>";
|
42| ?>

```

Hasilnya seperti di Gambar 2-3.

Resep di bawah untuk mengambil headline detikinet (<http://www.detikinet.com>) menggunakan regex. Karena situs ini bersaudaraan dengan detikcom, dan layoutnya kebetulan sama, kita dapat menggunakan pola regex yang persis sama untuk mengekstrak beritanya. Kali ini kita akan menampilkan juga paragraf summary yang kita ekstrak.



Gambar 2-3. Contoh Headline Detikcom.

```

1| <?
|
3| //
4| // headline-detikinet.php
5| //
|
7| if (($fp = fopen("http://www.detikinet.com/
  index.shtml", "r"))===false)
8|     die("Gagal mengambil URL!");
|
10| $content = "";
11| while ($chunk = fgets($fp)) $content .= $chunk;
|
13| if (!$content)
14|     die("Dokumen URL kosong!");
|
16| if (!preg_match_all('#<span
  class="tanggal(?:High)?">([>]+)</span><BR>' .
17|     '\s*<A href="([>]+)"[>]+<span
  class="(?:judulHigh|strJudul)?">' .

```

```

18|     '(.+?)</span></A><BR>\s*<span
  class="summary(?:High)?">' .
19|     '(.+?)</span>#si', $content, $mm, PREG_SET_ORDER))
20|     die("Gagal mengekstrak headline dari dokumen URL!");
|
22| $items = array();
23| foreach ($mm as $m) {
24|     $items[] = array(
25|         'date' => $m[1],
26|         'link' => $m[2],
27|         'title' => $m[3],
28|         'summary' => $m[4],
29|     );
30| }
|
32| echo "<h3>Headline <a href=\"http://www.detikinet.com\">".
33|     "detikInet</a></h3>";
|
35| foreach ($items as $item) {
36|     echo "<p><a href=$item[link]>$item[title]</a><br>";
37|     echo "$item[summary]</p>";
38| }
|
40| echo "</ul>";
|
42| ?>

```

Hasilnya seperti di Gambar 2-4.

Resep di bawah untuk mengambil headline kompas.co.id menggunakan regex.

```

1| #!/usr/bin/php -q
2| <?
|
4| //
5| // headline-kompas.php
6| //
|
8| $DB_HOST = "localhost";
9| $DB_DB = "test";
10| $DB_USER = "headlines";

```

```

11| $DB_PASS = "headline123";
12|
13|
14| if (($fp = fopen("http://www.detik.com/
    index.htm","r"))===false)
15|     die("Gagal mengambil URL!");
16|
17| $content = "";
18| while ($chunk = fgets($fp)) $content .= $chunk;
19|
20| if (!$content)
21|     die("Dokumen URL kosong!");
22|
23| # boks yang mengandung headline utama
24| if (!preg_match('#.<(table.+?<span class="sectionHL".+?</
    table>)#si',
25|         $content, $m))
26|     die("Gagal mengekstraksi boks utama dari dokumen URL!");
27| $boks_utama = $m[1];
28|
29| # ekstrak headline utama, tidak semua mengandung summary
30| if (!preg_match_all('#<a href="/[^[^"]+)" .+?><font[<^>]+>?'.
31|         '([<^>]+)</font>?</a>#si', $boks_utama, $mm,
32|         PREG_SET_ORDER))
33|     die("Gagal mengekstrak headline utama dari dokumen URL!");
34|
35| $items = array();
36| foreach ($mm as $m) {
37|     $items[] = array(
38|         'link' => "http://www.kompas.co.id/$m[1]",
39|         'title' => $m[3],
40|         'summary' => '',
41|     );
42| }
43|
44| # ekstrak headline sekunder, semua mengandung summary
45| if (!preg_match_all('#<span class="SectionHL2"><a href="/'
    ([^"]+)" '.
46|         'target="_parent">(.+?)</a></
    span>\s*<br>\s*(.+?)<br>#si',
47|         $content, $mm, PREG_SET_ORDER))

```

```

48| die("Gagal mengekstrak headline sekunder dari dokumen
    URL!");
49|
50| foreach ($mm as $m) {
51|     $items[] = array(
52|         'link' => "http://www.kompas.co.id/$m[1]",
53|         'title' => $m[2],
54|         'summary' => $m[3],
55|     );
56| }
57|
58| # ekstrak post_time dari URL
59| for ($i=0; $i<count($items); $i++) {
60|     if (preg_match("#/(\d\d)(\d\d)/(\d+)/(\d\d)(\d\d)(\d\d)#",
61|         $items[$i]['link'], $m)) {
62|         $m[1] += 2000; # buat tahun menjadi 4 digit
63|         $items[$i]['post_time'] =
            "$m[1]$m[2]$m[3]$m[4]$m[5]$m[6]";
64|     } else {
65|         $item[$i]['post_time'] = null;
66|     }
67| }
68|
69| # insert/replace ke database
70| if (!$db = mysql_connect($DB_HOST,$DB_USER,$DB_PASS))
71|     die("Gagal koneksi ke database!");
72| mysql_select_db($DB_DB);
73|
74| foreach ($items as $item) {
75|     mysql_query("REPLACE INTO headlines (
76|         insert_time,
77|         post_time,
78|         source,
79|         url,
80|         title,
81|         summary) VALUES (
82|             NOW(),".
83|             (is_null($item['post_time']) ?
                "NULL":"'$item[post_time]'",
84|             'KOMPAS',
85|             "'.mysql_escape_string($item['link']).'",

```

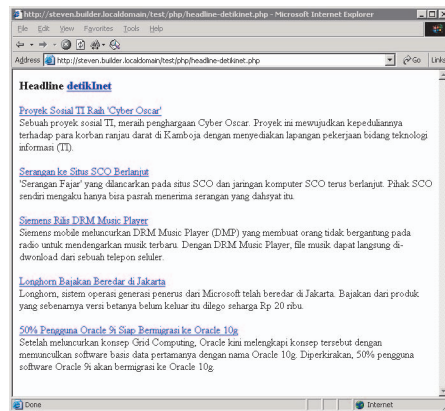
```

86|     '"mysql_escape_string($item['title'])."',
87|     '"mysql_escape_string($item['summary'])."'");
88| }
89|
90| ?>

```

Pembahasan

Dibandingkan situs detikcom atau detikInet, Kompas memiliki perbedaan dalam menyebutkan link. Link-link di situs ini relatif (), jadi kita perlu menambahkan "http://www.kompas.com/" agar menjadi absolut, baris 38 dan 52. Kemudian, kali ini kita tidak menampilkan ke browser melainkan menyimpannya di dalam database. Sehingga skrip kita jadikan skrip command-line saja. Kompas tidak menampilkan tanggal pada teks, namun URL berita mengandung informasi tanggal dan waktu sehingga kita dapat mem-parse-nya (baris 59-67).



Gambar 2-4. Headline dan Kutipan Berita.

Struktur tabel MySQL yang digunakan adalah sebagai berikut:

```

create table headlines (
  source varchar(10) not null,
  url varchar(255) not null, unique(url),
  title varchar(255) not null,
  summary varchar(255) not null,
  post_time datetime,
  insert_time datetime not null
);

```

Kolom source kita buat agar tabel dapat menyimpan headline berita dari sumber-sumber lain. Sebetulnya, agar lebih lengkap, kita perlu membuat juga tabel sumber headline, misalnya headline_sources dan membuat headlines.source sebagai foreign key dari headline_sources.id. Field insert_time untuk menyimpan kapan kita membuat record ini. Field post_time untuk mencatat kapan sebuah berita diterbitkan oleh sumber. Field ini dibuat NULL-able agar seandainya sebuah

sumber tidak menyediakan informasi waktu terbit di halaman URL-nya, bisa dibiarkan nilai field ini bernilai NULL atau waktu terbit berita tidak diketahui. Untuk memastikan tidak ada berita yang tersimpan ganda, kita membuat URL sebagai unique key dan melakukan insert dengan statement REPLACE INTO. Sehingga bila ada koreksi judul atau summary dari Kompas, kita dapat mengganti data lama di database dengan yang baru (baris 75-87). Sayangnya, tidak ada cara yang mudah untuk memastikan kita tidak kehilangan dan kelewatan sebuah berita. Anda perlu menjalankan skrip ini cukup sering atau membuat skrip yang dapat mengambil daftar headline dari halaman search atau arsip situs Kompas. Andaikata setiap berita memiliki ID yang berurutan (misalnya: 10010, 10011, 10012, dan seterusnya) tentu akan lebih mudah bagi mengambil headline secara berurutan tanpa kehilangan/kelewatan. Namun kebetulan situs Kompas—dan juga detik.com—menggunakan pola tanggal/bulan/tahun dan jam/menit/detik untuk memberi ID pada berita-beritanya.

Untuk menampilkan headline dari database, bisa menggunakan skrip di bawah ini:

```

1|#!/usr/bin/php -q
2|<?
3|
4|//
5|//  tampilkan-headline-kompas.php
6|//
7|
8|$DB_HOST = "localhost";
9|$DB_DB = "test";
10|$DB_USER = "root";
11|$DB_PASS = "pesatpesat";
12|
13|if (!$db = mysql_connect($DB_HOST,$DB_USER,$DB_PASS))
14|    die("Gagal koneksi ke database!");
15|mysql_select_db($DB_DB);
16|
17|$res = mysql_query("SELECT
18|    url,title,summary,
19|    DATE_FORMAT(post_time,'%d %b %Y @
20|    %H:%i')
21|    as post_time
22|    FROM headlines
23|    WHERE source='KOMPAS'
24|    ORDER BY insert_time DESC LIMIT 5");

```



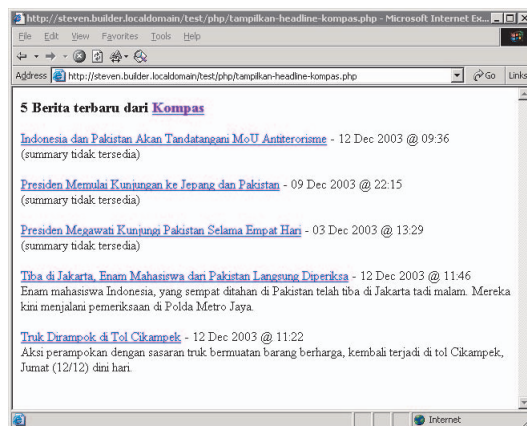
```

24| if (!$res) die("Gagal query dari database!");
25|
26| echo "<h3>5 Berita terbaru dari
27|      <a href=http://www.kompas.co.id>Kompas</a></h3>";
28|
29| while ($row = mysql_fetch_array($res)) {
30|     echo "<p><a href=$row[url]>$row[title]</a> - ",
31|         (is_null($row['post_time']) ?
32|         "(tanggal post tidak diketahui)" :
33|         $row['post_time']),
34|         "<br>",
35|         ($row['summary'] ? $row['summary'] : "(summary tidak
36|         tersedia)"),
37|         "</p>\n";
38| }
39| ?>

```

Hasilnya pada Gambar 2-5.

Mudah-mudahan beberapa contoh pada subbab ini bisa memberi gambaran bagaimana mengambil data dari situs lain. Seperti cuaca dari <http://meteo.bmg.go.id/prakiraan.asp> untuk kota-kota di Indonesia, info saham dari <http://www.indoexchange.com>, <http://www.e-bursa.com>, atau <http://www.detik.com>, dan masih banyak lainnya.



Gambar 2-5. Headline dari Database.

Resep 2-3: Mengambil Baris Acak dari File

```

1|<?php
2|
3|//
4|// baris-acak-memory.php
5|//
6|
7|// mengambil 1 baris acak dari file
8|function pick_random_line_1($file) {
9|    $lines = file($file);
10|    return $lines[rand(0, count($lines)-1)];
11|}
12|
13|// mengambil hingga N baris acak dari file
14|function pick_random_line_n($file, $N) {
15|    $lines = file($file);
16|    if ($N > count($lines)) $N=count($lines);
17|
18|    $picked = array();
19|    for ($i=0; $i<$N; $i++) {
20|        while (true) {
21|            $j = rand(0, count($lines));
22|            if ($lines[$j] != false) {
23|                $picked[] = $lines[$j];
24|                $lines[$j] = false;
25|                break;
26|            }
27|        }
28|    }
29|
30|    return $picked;
31|}
32|
33|?>

```

Fungsi `pick_random_line_1()` pada resep di atas adalah mengambil satu baris acak dari sebuah file. Fungsi `pick_random_line_n()` mengambil N buah baris

acak dari sebuah file atau hanya sebanyak jumlah baris pada file, jika N lebih besar dari jumlah baris file. Algoritma yang dipakai sangat sederhana, yaitu pertama menampung baris-baris file ke dalam array dengan fungsi `file()`, lalu memilih sebuah elemen acak dari array tersebut. Untuk fungsi `pick_random_line_n()`, proses pemilihan elemen diulangi sebanyak N kali, di mana elemen yang telah terpilih akan diset menjadi false agar tidak dipilih lagi.

Sebetulnya proses pemilihan 1 atau lebih elemen acak dari array dapat dilakukan dengan memanggil fungsi builtin PHP yaitu `array_rand()`.

Algoritma ini memiliki kelemahan yaitu boros memori. Kalau ukuran file amat besar, maka algoritma ini tidak bisa dipakai karena program akan kehabisan RAM. Alternatifnya, kita bisa melakukan 2 kali passing atau pembacaan file. Pass pertama untuk menghitung jumlah baris pada file, misalnya M. Lalu kita mengambil bilangan acak antara 1 hingga M. Baru pembacaan kedua untuk mengambil baris dengan nomor yang tadi dipilih.

Alternatif lainnya, terdapat algoritma untuk melakukan hal ini dalam 1 kali passing saja, sebagai berikut:

```

1|<?php
2|
3|//
4|// baris-acak-1pass.php
5|//
6|
7|// mengambil 1 baris acak dari file
8|function pick_random_line_1($file) {
9|    if (($fp = fopen($file, "r")) === false) return false;
10|    $i = 0;
11|    while ($line = fgets($fp, 8192)) {
12|        if (rand(0, $i) == 0) $picked = $line;
13|        $i++;
14|    }
15|    return $picked;
16|}
17|
18|// mengambil hingga N baris acak dari file
19|function pick_random_line_n($file, $N) {
20|    if (($fp = fopen($file, "r")) === false) return false;
21|    $i = 0;

```

```

22|    $picked = array();
23|    while ($line = fgets($fp, 8192)) {
24|        // belum semua terisi, isi dulu
25|        if (count($picked) < $N) {
26|            $picked[] = $line;
27|            $j1 = rand(0, count($picked)-1);
28|            $j2 = rand(0, count($picked)-1);
29|            $tmp=$picked[$j1]; $picked[$j1]=$picked[$j2];
30|            $picked[$j2]=$tmp;
31|        } else {
32|            $j=rand(0, count($picked)-1);
33|            $i++;
34|            if (rand(1, $i) <= $N) {
35|                $picked[$j] = $line;
36|                continue;
37|            }
38|        }
39|    }
40|    return $picked;
41|}
42|
43|?>

```

Mari kita periksa cara kerja algoritma ini dengan beberapa contoh. Perhatikan fungsi `pick_random_line_1()` pada kode, baris 11-13. Katakanlah kita mempunyai file 1 baris. Pada putaran loop `while()` pertama, `$i=0`, `rand(0, $i)` berpeluang 100% bernilai 0, `$picked` berpeluang 100% di-set ke isi baris 1. Setelah loop berakhir, peluang `$picked` berisi baris 1 adalah 100%. (Benar.)

Sekarang anggap kita mempunyai file 2 baris. Pada putaran pertama, `$i=0`, `rand(0, $i)` berpeluang 100% bernilai 0, `$picked` berpeluang 100% diset ke isi baris 1. Pada putaran kedua, `$i=1`, `rand(0, $i)` berpeluang 50% bernilai 0, `$picked` berpeluang 50% diset ke isi baris 2. Setelah loop berakhir, peluang `$picked` berisi baris 1 adalah $100\% \times 50\% = 50\%$. Peluang `$picked` berisi baris 2 adalah 50%. (Benar.)

Sekarang anggap kita mempunyai file 3 baris. Pada putaran pertama, `$i=0`, `rand(0, $i)` berpeluang 100% bernilai 0, `$picked` berpeluang 100% diset ke isi baris 1. Pada putaran kedua, `$i=1`, `rand(0, $i)` berpeluang 50% bernilai 0, `$picked` berpeluang 50% diset ke isi baris 2.

Pada putaran ketiga, $\$i=2$, $\text{rand}(0, \$i)$ berpeluang 33,3% bernilai 0, $\$picked$ berpeluang 33,3% diset ke isi baris 3 (dengan kata lain, $\$picked$ berpeluang 66,7% tetap tidak berubah dari nilai putaran sebelumnya). Setelah loop berakhir, peluang $\$picked$ berisi baris 1 adalah $100\% \times 50\% \times (1-33,3\%) = 33,3\%$. Peluang $\$picked$ berisi baris 2 adalah $50\% \times (1-33,3\%) = 33,3\%$. Peluang $\$picked$ berisi baris 3 adalah 33,3%. (Benar.)

Fungsi `pick_random_line_n()` prinsip kerjanya sama, namun di baris 25-30 akan selalu mengisi dulu $\$picked$ hingga jumlah elemen mencukupi target N buah.

Resep 2-4: Mengambil Data dari File CSV

File CSV, *comma-separated value*, merupakan jalur yang termudah dalam mengeksport data dari spreadsheet seperti Microsoft Excel ke dalam database MySQL atau ke dalam struktur data PHP.

Format CSV terdiri dari satu baris atau lebih, di mana setiap baris merupakan sebuah record. Setiap record terdiri dari satu field atau lebih yang dipisahkan dengan tanda koma (,). Kalau ada nilai field yang mengandung koma, maka field ini dapat dikutip dengan tanda kutip ganda (") sehingga tanda koma tidak akan diinterpretasi sebagai pemisah field. Tanda pemisah field dan kutip dapat berbeda-beda.

Contohnya, jika komputer Windows Anda di-set pada locale Indonesia, maka Excel akan mengimpor CSV dengan pemisah field berupa titik koma (;) bukan koma. Ini karena koma banyak digunakan sebagai pemisah koma desimal untuk locale Indonesia. Sementara jika komputer Windows Anda diset pada locale Amerika atau Inggris, Excel akan memilih menggunakan koma sebagai pemisah field karena untuk pemisah desimal menggunakan titik (.).

MySQL memiliki statement `LOAD DATA INFILE` yang mengizinkan kita menyebutkan pemisah field dan tanda kutip, sehingga MySQL dengan mudah dapat mengimpor file CSV. Bagaimana dengan PHP?

Jika sebuah file CSV tidak pernah mengandung tanda kutip, maka kita bisa mem-parse setiap baris atau record dengan `preg_split(" ", $baris)`. Namun jika mengandung kutip, maka regex biasa tidak bisa mem-parse-nya dengan benar. Untuk itu PHP menyediakan fungsi `fgetcsv()` yang bisa mem-parse kolom-kolom

field dengan benar, dan mengizinkan kita mengatur spesifikasi karakter pemisah field dan tanda kutip yang diinginkan.

Berikut ini contoh resep mem-parse sebuah file CSV:

```
1|<?
2|
3|//
4|// csv-1.php
5|//
6|
7|$FIELD_SEPARATOR = ",";
8|$FIELD_QUOTE = '"';
9|$SKIP_HEADER_ROW = true;
10|$CSV_FILE = "resephp.csv";
11|
12|function fgetcsv($fp) {
13|    global $FIELD_SEPARATOR, $FIELD_QUOTE;
14|
15|    if (version_compare(PHP_VERSION, "4.3.0") >= 0)
16|        return fgetcsv($fp, 8192, $FIELD_SEPARATOR,
17|            $FIELD_QUOTE);
18|    else
19|        return fgetcsv($fp, 8192, $FIELD_SEPARATOR);
20|}
21|
22|$fp = fopen($CSV_FILE, "r");
23|
24|$i=0;
25|$rows = array();
26|while ($row = fgetcsv($fp)) {
27|    $i++;
28|    if ($SKIP_HEADER_ROW && $i==1) continue;
29|    $rows[] = $row;
30|}
31|
32|// gunakan hasilnya
33|echo "<pre>";
34|print_r($rows);
35|echo "</pre>";
36|?>
```

Perhatikan contoh file `resepphp.csv` di halaman berikut ini:

```
nama_resep,id_resep,id_kategori,nama_kategori,kata_kunci
Mengambil Kurs,2-1,2,Memperoleh Data,kurs; dolar; rupiah
Mengambil Headline Berita,2-2,2,Memperoleh Data,berita; headline
berita
Mengambil Baris Acak Dari File,2-3,2,Memperoleh Data,random;
acak; rand()
Mengambil Data dari File CSV,2-4,2,Memperoleh Data,CSV;
fgetcsv()
Konversi Tanggal/Waktu,3-1,3,Memanipulasi dan Mengkonversi
Data,tanggal; waktu
Konversi Zona Waktu,3-2,3,Memanipulasi dan Mengkonversi
Data,zona waktu; timezone
```

Hasilnya akan ditampung dalam array of array PHP:

```
Array
(
    [0] => Array
        (
            [0] => Mengambil Kurs
            [1] => 2-1
            [2] => 2
            [3] => Memperoleh Data
            [4] => kurs; dolar; rupiah
        )
    [1] => Array
        (
            [0] => Mengambil Headline Berita
            [1] => 2-2
            [2] => 2
            [3] => Memperoleh Data
            [4] => berita; headline berita
        )
    [2] => Array
        (
            [0] => Mengambil Baris Acak Dari File
            [1] => 2-3
```

```
                [2] => 2
                [3] => Memperoleh Data
                [4] => random; acak; rand()
            )
    [3] => Array
        (
            [0] => Mengambil Data dari File CSV
            [1] => 2-4
            [2] => 2
            [3] => Memperoleh Data
            [4] => CSV; fgetcsv()
        )
    [4] => Array
        (
            [0] => Konversi Tanggal/Waktu
            [1] => 3-1
            [2] => 3
            [3] => Memanipulasi dan Mengkonversi Data
            [4] => tanggal; waktu
        )
    [5] => Array
        (
            [0] => Konversi Zona Waktu
            [1] => 3-2
            [2] => 3
            [3] => Memanipulasi dan Mengkonversi Data
            [4] => zona waktu; timezone
        )
)
```

Pembahasan

Fungsi `fgetcsv()` digunakan untuk membaca sebuah baris record. Fungsi ini memiliki empat parameter. Parameter pertama dan kedua sama seperti pada `fgets()`, yaitu file handle dan ukuran baris maksimum. Parameter ketiga adalah pemisah field, parameter keempat adalah karakter kutip. Di baris 12-19 ada fungsi *wrapper* untuk `fgetcsv()` karena pada versi sebelum 4.3.0, `fgetcsv()` belum mendukung parameter keempat. Karena itu, abaikan parameter keempat pada versi PHP yang lebih awal.

Resep 2-4: Mengambil Data dari File CSV

Perlu dicatat, Microsoft Excel juga dapat menghasilkan file CSV dengan satu record terdiri dari beberapa baris, yaitu jika sebuah field mengandung newline. Jadi tanda kutip akan mengutip nilai field yang berlaku dari satu baris ke baris berikutnya. File-file seperti ini tidak dapat di-parse dengan kode contoh pada resep ini.