

How to Download Manga from Manga Stream (After They Split the Image)

Pendahuluan (Ngemeng section)

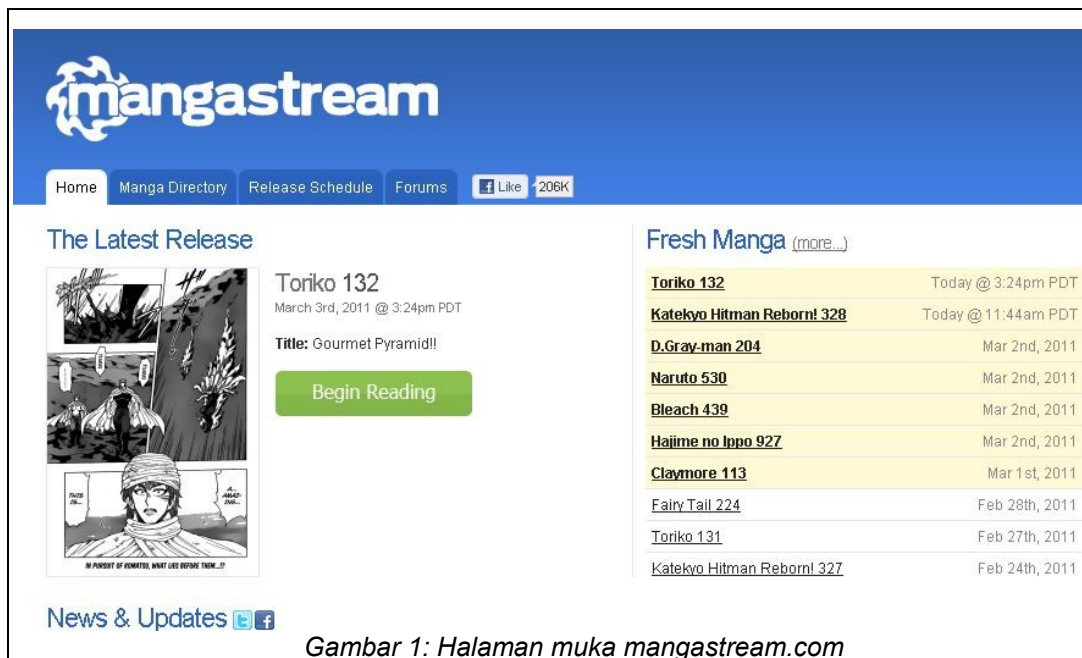
One Piece adalah manga paling keren (arguably, but not with me) di Internet saat ini. Tim mangascan One Piece yang paling aktif dan paling banyak diikuti saat ini adalah Manga Stream dengan website mangastream.com. Karena mangastream.com bukan satu-satunya website online manga reader, ada yang mengembangkan ekstensi Google Chrome All Manga Reader yang memudahkan user menikmati website online manga reader. Normalnya, user melihat 1 gambar tiap 1 halaman, ekstensi ini bisa membuat 1 halaman berisi semua gambar 1 chapter.

Ternyata, MangaStream tidak suka dengan keberadaan benda seperti ini sehingga mereka melakukan trik untuk menggagalkannya. Alhasil, sejak MangaStream merilis Claymore 113, 1 halaman memang berisi 1 gambar, tetapi gambar ini terdiri atas beberapa potongan gambar. Akibatnya, kita tidak bisa lagi klik-kanan-save-as per halaman untuk menyimpan manga dari mangastream.com, harus menyatukan potongan2 di satu halaman atau printscreen dan crop per halaman.

Saya, sebagai orang yang malas, sudah tentu terlalu malas untuk melakukan hal ini, jadilah saya membuat script PHP (coz I like PHP) sederhana yang secara otomatis mendownloadkan satu chapter gambar dari mangastream, menyatukan potongan-potongan gambar, dan me-rename masing-masing gambar agar lebih tertata. Pada (self-proclaimed) paper ini saya akan menjelaskan teknik untuk mengotomatisasikan proses download manga dari situs mangastream.com dengan bantuan bahasa PHP (logikanya bisa diaplikasikan ke bahasa apapun (bukan, bukan bahasa Sunda, Cina, Jawa, ataupun bahasa manusia lainnya)).

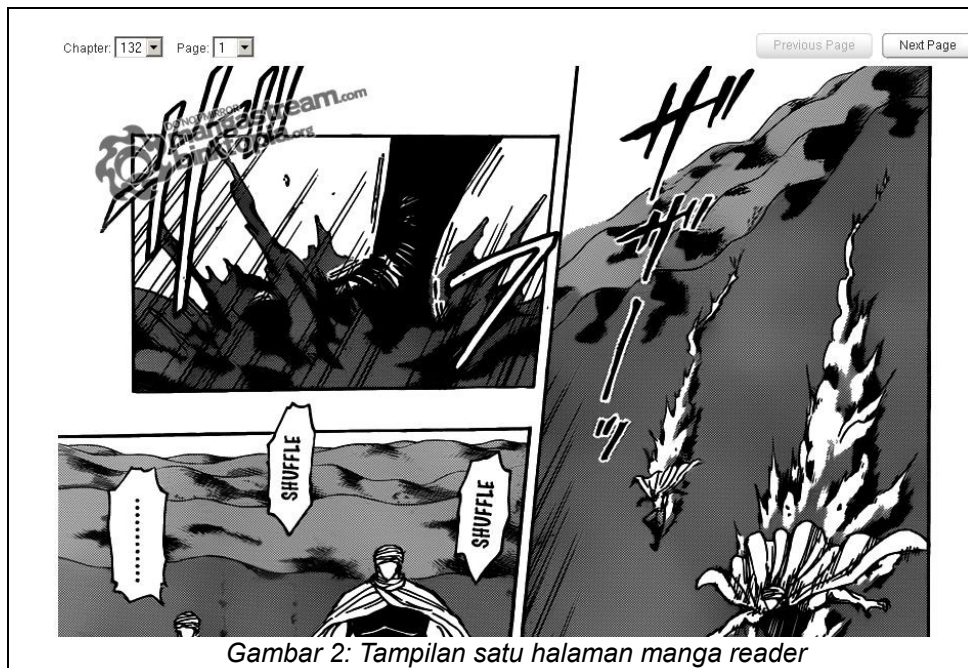
Pengenalan Mangastream.com (More Ngemeng Section)

Mangastream.com adalah situs milik tim mangascan Manga Stream join project Binktopia. Situs ini paling terkenal karena berisi mangascan One Piece paling cepat dan aktif saat ini.



Gambar 1: Halaman muka mangastream.com

Seperti beberapa situs tim mangascan, situs ini pun memiliki online manga reader sendiri dan buatan sendiri (ada juga yang pakai foolreader atau OMV). Normalnya, satu halaman manga reader berisi satu gambar (satu halaman komik).



Gambar 2: Tampilan satu halaman manga reader

Namun, ada yang membuat script, baik itu berupa extension browser ataupun Greasemonkey script, untuk memudahkan pembaca, yaitu satu halaman berisi gambar satu chapter. Selain itu, bagi pembaca yang ingin mengoleksi (men-download) manga tertentu (di mana tidak disediakan link direct download), dia bisa melakukan kombo klik-kanan-save-as pada tiap-tiap gambar.

Nampaknya, mangastream.com tidak suka akan hal ini, entah tidak suka kepada extension semacam All Manga Reader atau kepada pembaca yang men-download gambar, sehingga mereka memutuskan untuk mempersulit hal itu dengan memotong-motong gambar pada tiap halaman. Ya, sejak me-release Claymore 113, mereka membuat satu halaman manga reader berisi beberapa potongan gambar yang membentuk satu gambar.



Gambar 3: Satu gambar terdiri atas beberapa potongan

Perubahan ini tidak berpengaruh bagi pembaca yang sekedar membaca online, tetapi sangat menyusahakan para pemakai kombo klik-kanan-save-as karena hanya memperoleh satu segmen per kombo dan bagi para penulis script karena mereka harus menulis ulang script mereka.

Investigasi Terhadap Online Manga Reader Mangastream.com (Beware: Source Code Ahead)

Mari kita lihat bagaimana sang manga reader menyusun potongan-potongan tersebut. Jika kita melakukan kombo Ctrl-U-view-source pada satu halaman, akan terlihat bahwa pada bagian headstyle terdapat kode CSS seperti ini yang memberitahu posisi tiap-tiap potongan:

```
#pebd5b24d671f57eee43c3302abea883f {position:absolute;z-index:4623;width:188px;height:124px;top:752px;left:0px}
#pf09f8fba09341725d777a4c456c36f0d {position:absolute;z-index:4549;width:426px;height:554px;top:0px;left:0px}
#p677b010c75ac9ec3760939c702e23579 {position:absolute;z-index:800;width:931px;height:424px;top:876px;left:0px}
#p871c5786201c628bd4b730995bbbbb1c6 {position:absolute;z-index:1495;width:505px;height:554px;top:0px;left:426px}
#pab38089d26f4f298962e3e3c793a4006 {position:absolute;z-index:7226;width:743px;height:124px;top:752px;left:188px}
#p0d560c82bb8885482394061d5f30d483 {position:absolute;z-index:770;width:243px;height:109px;top:163px;left:73px}
#pba795e6ce94394b5d05571d05afcd370 {position:absolute;z-index:4416;width:737px;height:198px;top:554px;left:194px}
#p6436c4d450b563bac0ae270019ae41d0 {position:absolute;z-index:1733;width:194px;height:198px;top:554px;left:0px}
```

Code 1: Cuplikan kode headstyle manga reader

Dan pada body HTML terdapat kode yang berisi URL gambar tiap-tiap potongan:

```
<div style="position:relative;width:931px;height:1300px">
  <div id="pebd5b24d671f57eee43c3302abea883f"><a href="/read/billy_bat/18964888/2"></a></div>
  <div id="pf09f8fba09341725d777a4c456c36f0d"><a href="/read/billy_bat/18964888/2"></a></div>
  <div id="p677b010c75ac9ec3760939c702e23579"><a href="/read/billy_bat/18964888/2"></a></div>
  <div id="p871c5786201c628bd4b730995bbbbb1c6"><a href="/read/billy_bat/18964888/2"></a></div>
  <div id="pab38089d26f4f298962e3e3c793a4006"><a href="/read/billy_bat/18964888/2"></a></div>
  <div id="p0d560c82bb8885482394061d5f30d483"><a href="/read/billy_bat/18964888/2"></a></div>
  <div id="pba795e6ce94394b5d05571d05afcd370"><a href="/read/billy_bat/18964888/2"></a></div>
  <div id="p6436c4d450b563bac0ae270019ae41d0"><a href="/read/billy_bat/18964888/2"></a></div>
</div>
```

Code 2: Cuplikan kode body manga reader

Dengan menggabungkan kedua informasi ini, kita dapat membuat script yang akan otomatis melakukan kombo buat-canvas-download-semua-potongan-tempel-ke-kanvas-export-jadi-satu-gambar.

Rancangan Solusi (Ini Dia)

Solusi yang saya pakai pada paper ini adalah menggunakan bahasa PHP dan ekstensi GD untuk mendownload dan menggabungkan potongan-potongan menjadi satu gambar utuh.

Alat dan Bahan (Emangnya Praktikum?)

- Editor Notepad++ (kalo ga pake notepad++ ya silahkan)
- Interpreter PHP 5.1.6 ke atas
- Ekstensi PHP GD
- Koneksi Internet
- PHP Class Crawler buatan captain_kuro <http://pastebin.com/fHe7Wwys>

Crawler adalah class PHP yang saya buat untuk membantu saya melakukan perampokan crawling situs-situs tertentu. Crawler beroperasi line-based yang artinya memproses one-line-at-a-time. Logika pemakaiannya seperti "pergi ke baris yg berisi ini; ambil bagian ini dari baris sekarang; pergi ke baris yg berisi dst...".

Algoritma (Bahasa Manusianya)

Dengan menggunakan class Crawler di atas, cara saya berpikir adalah seperti ini:

```
// Gather Information
1 Buka satu halaman manga reader (misal http://mangastream.com/read/billy_bat/18964888/1)
2 Pergi ke baris yang berisi definisi CSS salah satu potongan
  (regex '/#.+position.+width.+height.+top.+left/')
3 Iterasi hingga ketemu baris penutup (berisi '-->')
  3a Ambil informasi id, z-index, height, width, left, top tiap potongan
    (regex '/#(\\w+) .+width:(\\d+).*height:(\\d+).*top:(\\d+).*left:(\\d+)/')
  3b Masukkan ke array (var $imgs)
4 Pergi ke baris yang berisi ukuran total gambar
  (misal <div style="position:relative;width:907px;height:1300px">)
5 Ambil $tot_width dan $tot_height dari baris ini
  (regex '/<div .+position:relative.+width:(\\d+).*height:(\\d+)/')
6 Iterasi hingga ketemu baris penutup (regex '/^\\s+<\\div>/')
  6a Ambil informasi id, src tiap potongan
    (regex '/<div id="(\\^")+".+src="(\\^")+"/')
  6b Gabungkan ke array tadi (var $imgs)
7 Setelah seluruh informasi potongan didapat, urutkan ascending berdasarkan z-index
  (untuk mengatasi potongan pengganggu, lihat Gambar 4)

// Generate Image
8 Create canvas berukuran $tot_iwidth x $tot_height (function imagecreatetruecolor())
9 Iterasi array berisi informasi tadi
  9a Download potongan (function file_put_contents() dan file_get_contents())
  9b Convert jadi canvas (function imagecreatefromjpeg/png())
  9c Tempelkan ke canvas utama (function image_copy())
  9d Hapus potongan (function unlink())
10 Export menjadi jpg/png
```

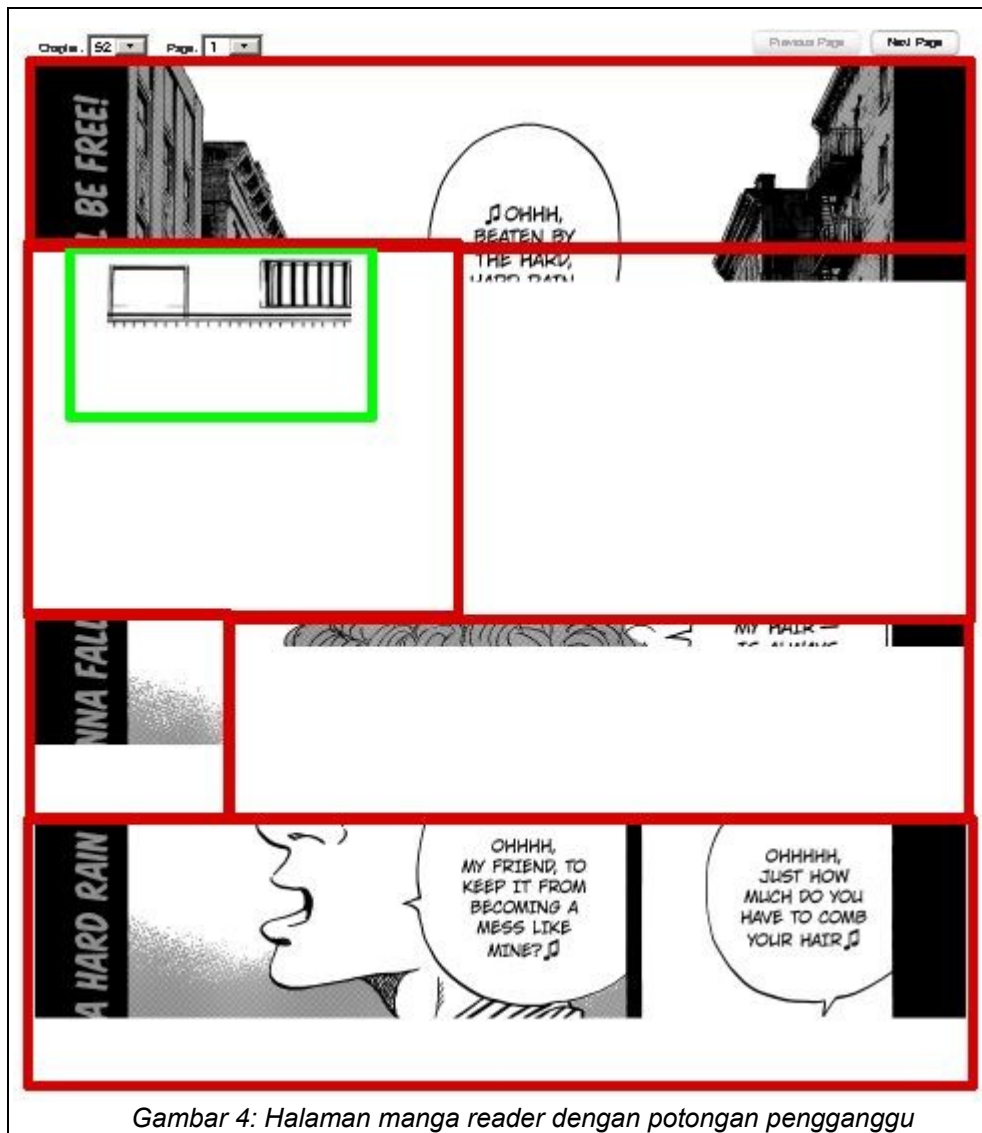
Code 3: Algoritma solusi

Tidak serumit yang dibayangkan kan? :|

Sedikit membahas potongan CSS di atas (Code 1), terlihat bahwa terdapat 6 atribut di mana yang penting bagi kita adalah 5 (mengabaikan position) yaitu width, height, top, left, dan z-index. Width x height menjadi ukuran potongan dan (left, top) menjadi koordinat posisi potongan pada gambar besar. Nah, apa itu z-index? Bayangkan seandainya ada elemen HTML yang overlapping, bagaimana browser menentukan elemen ini ditaruh di depan elemen itu? Well, ternyata CSS mendefinisikan “dimensi ketiga” dengan “sumbu Z” pada atribut z-index ini. Defaultnya semua elemen memiliki z-index yang sama, makin besar z-index, makin ada di depan.

Apa hubungan z-index dengan manga reader? Kadang ada potongan pengganggu pada halaman reader, contohnya Gambar 4.

Ketika seluruh potongan selesai di-download browser, memang potongan pengganggu tidak tampak karena memiliki z-index yang lebih kecil dibandingkan potongan lain, tetapi jika kita membuat script otomatis yang mengabaikan z-index, ada kemungkinan potongan pengganggu akan muncul pada gambar gabungan. Untuk itulah dilakukan langkah 7 (Code 3).



Gambar 4: Halaman manga reader dengan potongan pengganggu

Source Code (Langsung Copas Ga Dijamin Jalan)

Ini adalah source code yang digunakan untuk mendownload menggabungkan potongan dalam satu halaman menjadi satu gambar. Anda dapat memodifikasinya agar men-download satu chapter sekaligus.

```
<?php
require 'crawler.php';

$url = 'http://mangastream.com/read/billy_bat/18964888/1'; // The page url
$dir = 'd:/temp/'; // Where to store
$filename = 'test'; // Rename to this

// 1 Buka satu halaman manga reader
$c = new Crawler($url);
$imgs = array();
// 2 Pergi ke baris yang berisi definisi CSS salah satu potongan
$c->go_to('/#.+position.+width.+height.+top.+left/', '', true);
// 3 Iterasi hingga ketemu baris penutup (berisi '-->')
while ($line = $c->readline()) {
    if (preg_match('/#(\\w+) .+width:(\\d+) .+height:(\\d+) .+top:(\\d+) .+left:(\\d+)/', $line,
    $match)) {
        // 3a Ambil informasi id, z-index, height, width, left, top tiap potongan
        list($all, $id, $width, $height, $top, $left) = $match;
        if (preg_match('/z-index:(\\d+)/', $line, $match)) $zindex = $match[1]; else $zindex
        = 0;
        // 3b Masukkan ke array (var $imgs)
        $imgs[$id] = array('id'=>$id, 'zindex'=>$zindex, 'width'=>$width, 'height'=>$height,
        'top'=>$top, 'left'=>$left);
    } else if (Crawler::is_there($line, '-->')) {
        break;
    }
}
// 4 Pergi ke baris yang berisi ukuran total gambar
```



```

$reg = '/<div .+position:relative.+width:(\\d+).+height:(\\d+)/';
$c->go_to($reg, '', true);
preg_match($reg, $c->curline, $match);
// 5 Ambil $tot_width dan $tot_height dari baris ini
list($all, $tot_width, $tot_height) = $match;
// 6 Iterasi hingga ketemu baris penutup (regex '/^\\s+<\\div>/')
while ($line = $c->readline()) {
    if (preg_match('/<div id="([^\"]+)" .+src="([^\"]+)"', $line, $match)) {
        // 6a Ambil informasi id, src tiap potongan
        list($all, $id, $src) = $match;
        // 6b Gabungkan ke array tadi (var $imgs)
        $imgs[$id]['src'] = $src;
        $imgs[$id]['filename'] = basename($src);
        $imgs[$id]['ext'] = strtolower(Crawler::cutfromlast1(basename($src), '.'));
    } else if (preg_match('/^\\s+<\\div>/', $line)) {
        break;
    }
}
// 7 Setelah seluruh informasi potongan didapat, urutkan ascending berdasarkan z-index
function the_comp($a, $b) {
    if ($a['zindex'] == $b['zindex']) return 0;
    return ((int)$a['zindex'] < (int)$b['zindex']) ? -1 : 1;
}
usort($imgs, 'the_comp');
// 8 Create canvas berukuran $tot_iwidth x $tot_height
$canvas = imagecreatetruecolor($tot_width, $tot_height);
// 9 Iterasi array berisi informasi tadi
foreach ($imgs as $img) {
    // 9a Download potongan
    file_put_contents($dir . $img['filename'], file_get_contents($img['src']));
    // 9b Convert jadi canvas
    if ($img['ext'] == 'png') {
        $tmpimg = imagecreatefrompng($dir . $img['filename']);
    } else {
        $tmpimg = imagecreatefromjpeg($dir . $img['filename']);
    }
    // 9c Tempelkan ke canvas utama
    imagecopy($canvas, $tmpimg,
        $img['left'], $img['top'], // koordinat di canvas
        0, 0, // koordinat ambil dari potongan
        $img['width'], $img['height'] // besar potongan yg di-copy
    );
    // 9d Hapus potongan
    unlink($dir . $img['filename']);
}
// 10 Export menjadi jpg/png
$ext = $imgs[count($imgs)-1]['ext'];
$fullpath = "$dir$filename.$ext";
if ($ext == 'jpg') {
    imagejpeg($canvas, $fullpath, 90);
} else {
    imagepng($canvas, $fullpath, 9, PNG_NO_FILTER);
}

```

Fyuuuh, what a code! Very dirty! :) Sayangnya hingga saat ini belum ada dokumentasi untuk class Crawler yang dipakai pada kode di atas (sang pembuat terlalu malas untuk menulis dokumentasi toh dipake sendiri). Untuk mencoba script di atas, letakkan file `crawler.php` dan `kode_di_atas.php` pada direktori yang sama, ganti nilai `$url` dengan halaman yang diinginkan, edit nilai `$dir` ke tempat di mana gambar gabungan ingin disimpan, dan ubah nilai `$filename` ke nama file yang diinginkan untuk satu gambar ini.

Satu hal lagi yang ingin saya sampaikan, yaitu tentang image quality ketika export. Berhubung selama ini mangastream.com hanya berisi gambar `.jpg` dan `.png` maka saya hanya menggunakan fungsi `imagecreatefrompng()` dan `imagecreatefromjpeg()`. Dan untuk mempertahankan size dan quality, saya pun men-export balik image gabungan sesuai format potongannya menggunakan `imagejpeg()` dan `imagepng()`.

Fungsi `imagejpeg()` akan men-export canvas menjadi file JPEG fisik. Fungsi ini menerima parameter tambahan berupa quality (parameter ketiga) dalam rentang 1-100 (default 75). Setelah melakukan beberapa eksperimen, saya merasa 90 adalah angka yang optimal (100 terlalu besar dan 75 memiliki penurunan kualitas yang “agak terlihat”).

Fungsi `imagepng()` akan men-export canvas menjadi file PNG fisik. Fungsi ini menerima dua parameter tambahan berupa compression level (parameter ketiga) dalam rentang 0-9 dan filter-filter yang digunakan (parameter keempat) dalam bentuk konstanta yang lengkapnya dapat dilihat pada PHP manual. Setelah bereksperimen dengan berbagai setting, saya mendapatkan bahwa compression 9 (max) dan no filter (`PNG_NO_FILTER`) adalah settingan dengan hasil terbaik dalam ukuran terkecil.

Referensi (Daftar link)

- [1] <http://www.mangastream.com/>
- [2] <https://chrome.google.com/extensions/detail/gcamgngbdaiibdbikdfmjinnmgoppecp>
- [3] <http://id2.php.net/manual/en/>
- [4] <http://captainkuro.com>
- [5] <http://captainkuro.byethost3.com>

Tentang penulis (Narsis section)

Kunjungi saja <http://captainkuro.byethost3.com> dan <http://captainkuro.com>

