

# Memproses Dokumen Microsoft Word di Web

Bagi Anda yang terbiasa bekerja dengan OpenOffice. Ada kalanya, ketika bepergian dan membawa dokumen OpenOffice dan Anda harus melihat isinya sementara tidak ada OpenOffice yang terinstal di komputer yang Anda gunakan (ketika di warnet misalnya), apa yang harus dilakukan? Pasti akan lebih mudah apabila Anda memanfaatkan program untuk meng-*convert* dokumen word tersebut ke format yang bisa dibaca di komputer yang sedang digunakan. Di artikel ini, kita akan membahas pembuatan program tersebut.

**M**asalah lupa meng-*convert* dokumen ke format yang lebih umum cukup sering penulis alami. Ketika bepergian dengan satu dokumen dan terpaksa harus melihat isinya sementara komputer yang sedang digunakan tidak terinstal OpenOffice misalnya, merupakan masalah yang cukup mengganggu. Kita sebenarnya bisa saja mengextract dokumen SXW menggunakan WinZip, lantas membaca file XML yang terkandung di dalamnya. Bisa saja. Tapi, bagaimana kalau tidak ada WinZip juga? Seperti yang telah dikatakan, akan lebih mudah apabila memanfaatkan *converter online* untuk dokumen word. Anda bisa memilih untuk meng-convert ke format yang lebih umum seperti TEXT atau PDF.

Di dalam artikel kali ini, kita akan memanfaatkan PHP, antiword dan program ps2pdf untuk membangun converter online dokumen Word ke beberapa format lain seperti TEXT, PDF ataupun PS. Aplikasi yang kita bangun dapat dihost di server web untuk dapat pula menyediakan servis gratis kepada pengguna internet lain.

Tentunya, hasil konversi tidaklah sempurna. Program yang kita gunakan untuk meng-convert, antiword, secara umum hanya dapat mengekstrak teks dan atribut-atribut *layout* sederhana lainnya. Gambar terkadang bisa, namun tidak bisa diharapkan untuk mirip seperti dokumen word aslinya. Tapi, tentunya tidak masalah. Dapat melihat isi saja sudah sangat berguna. Anda bahkan bisa melihat *hidden text* dokumen MS Word Anda.

Artikel ini ditulis di distro SUSE Pro 9.1, PHP 4.3.4, Antiword 0.35 dan Ghostscript library 7.07. Untuk distro lain, harusnya tidak banyak hal yang berubah apabila semua program yang dibutuhkan telah terinstal.

## Mengenal Antiword

Sebelum kita melangkah ke aplikasi, ada baiknya apabila kita mengenal program antiword terlebih dahulu. Program Antiword adalah program yang ditujukan untuk menampilkan isi dari dokumen MS Word. Pada awalnya, format yang digunakan untuk output hanyalah file teks. Namun seiring dengan perkembangannya, saat ini, program ini juga bisa dapat meng-convert ke format PS. Format *output* teks tentunya memiliki sejumlah keterbatasan seperti gambar dan atribut layout lain. Dengan format PS, hal-hal tersebut bisa ditangani. Bagi yang XML mania, Anda juga bisa memanfaatkan antiword untuk meng-convert ke format XML.

Program ini bisa di-*download* di <http://www.winfield.demon.nl/>, walaupun, umumnya antiword sudah dipaketkan bersama distro Anda. Cobalah untuk melihat ke dalam CD/DVD distribusi Anda sebelum men-*download* dan mengompilasi dari source.

Cara menggunakan program ini sangatlah mudah. Anda cukup memberikan argumen berupa nama file dokumen MS Word, dan secara default, antiword akan meng-convert dokumen tersebut ke format TEXT

dan menampilkannya ke standard output (layer). Contoh:

```
$ antiword a.doc
```

Untuk meng-convert ke format XML, Anda bisa memberikan opsi -x diikuti oleh nama DTD. Antiword yang penulis gunakan hanya bisa menggunakan DTD berupa db(docbook). Berikut ini adalah contoh konversi dokumen MS Word ke format XML:

```
$ antiword -x db a.doc
```

Untuk meng-convert ke format PS, Anda bisa memberikan opsi -p, diikuti oleh ukuran kertas seperti letter atau a4. Ukuran kertas diberikan dalam huruf kecil. Dalam beberapa hal, apabila Anda mempergunakan encoding UTF-8, kombinasi Postscript dan UTF-8 tidaklah didukung. Oleh karena itu, kita perlu memberikan opsi -m, untuk memberikan file yang berisikan character-mapping agar file PS tetap bisa dihasilkan. File-file character mapping tersebut sudah datang bersama paket antiword dan umumnya terletak di `/usr/share/antiword`. Anda harus menyebutkan nama filenya saja dan bukan path lengkap untuk opsi -m. Contoh berikut ini akan mengconvert a.doc format PS (ukuran kertas letter) dengan menggunakan pemetaan karakter UTF-8 ke ISO-8859-1:

```
$ antiword -pletter -m8859-1.txt  
a.doc > a.ps
```

Program ini akan selalu mencetak hasil konversi ke standard output. Apabila kita

ingin menyimpan ke dalam file, kita harus meredireksi standard output ke nama file seperti contoh sebelumnya.

Catatan tambahan untuk konversi ke format PS. Antiword memiliki dua opsi tambahan yaitu `-i` untuk image level dan `-L` untuk mode landscape. Gunakan sesuai preferensi Anda.

Baik. Antiword mampu meng-convert ke PS dan XML. Tapi saya ingin format PDF. Apa yang harus saya lakukan? Antiword memang tidak bisa mengconvert langsung ke format PDF, namun Antiword sudah melakukan pekerjaan yang sangat baik dengan dapat mengconvert ke format TEXT dan format PS, sebab:

- Anda bisa meng-convert langsung format TEXT ke format PDF dengan memanfaatkan program `text2pdf` yang bisa didapatkan di <http://www.eprg.org/pd-corner/text2pdf/>. Program ini mampu meng-convert format TEXT ke PDF 1.1 dengan sangat cepat dan cukup bersih. Dalam tulisan kali ini, kita tidak memanfaatkan program ini karena setelah beberapa kali mencoba, penulis lebih senang untuk menggunakan cara yang akan kita bahas selanjutnya.
- Anda bisa memanfaatkan output PS dari antiword dan menggunakan program `ps2pdf` yang datang bersama paket `ghostscript` library. Dengan demikian, kita memberi perintah sama seperti ketika kita ingin meng-convert ke format PS. Setelah itu, kita melewati output antiword ke program `ps2pdf` dan menyimpan hasil konversi ke sebuah file PDF.

Sebagai contoh untuk konversi ke PDF memanfaatkan `ps2pdf`:

```
$ antiword -m8859-1.txt -pletter
a.doc | ps2pdf - > a.pdf
```

Demikianlah pengenalan menggunakan program antiword. Berikut ini, kita akan melihat bagaimana memanfaatkan program-program tersebut untuk menghasilkan konverter online dokumen MS Word.

### Doconweb

Secara umum, aplikasi yang akan kita bangun sangatlah sederhana. Apa yang kita butuhkan adalah file yang di-upload oleh user. Setelah file diterima, kita hanya perlu memanggil

program antiword dan `ps2pdf` (apabila dibutuhkan) untuk menghasilkan output sesuai format output yang dipilih oleh user. Output tersebut kemudian dikembalikan ke browser lengkap dengan mime type-nya dan browser akan menangani output tersebut sesuai pengaturan pada browser.

Anda bisa saja membangun service untuk publik. Masalahnya adalah, siapa yang mau memberikan (meng-upload) dokumennya begitu saja kepada pihak yang tidak dikenal.

Berikut ini adalah source code `doconweb.php`. Penjelasan setelah source code.

```
<?
/*
doconweb v0.1
nop, 10052005, 10:48
gpl
*/

$app_name = "doconweb";
$app_version = "0.1";
$app_copy = "(c) nop 2005, GPL";
$app_extra = "(use ghostscript
library (c) Peter Deutsch
<ghost@aladdin.com> and
antiword (c) A.J. van Os
<antiword@winfield.demon.nl>)";

$converter_app = "/usr/bin/
antiword";
$mapping_file = "8859-1.txt";

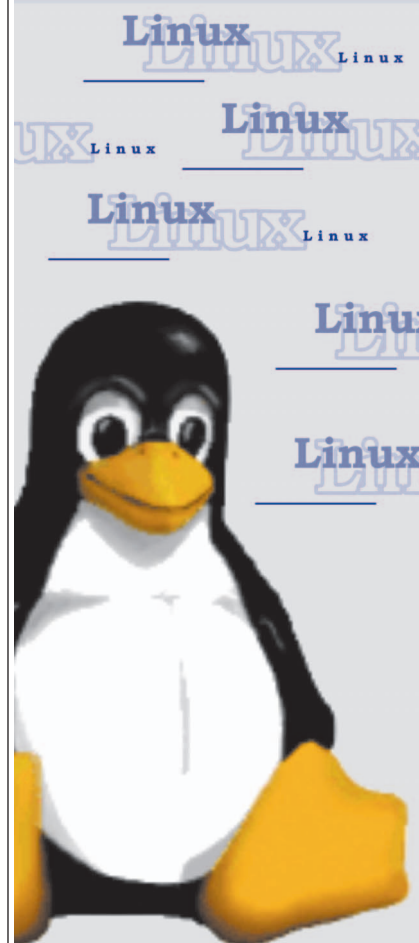
if (!$_POST)
{
echo "
<html>
<head>
<title></title>
</head>
<body>
<h3>$app_name $app_version</
h3> $app_extra
<hr noshade>
";
echo "
<form action='{$_
SERVER['PHP_SELF']}' method=
'post' enctype='multipart/form-
data'>
<table border=1>
<tr><td>Word
```

Open Source:

## GudangLinux

freedom forever

<http://www.gudanglinux.com>

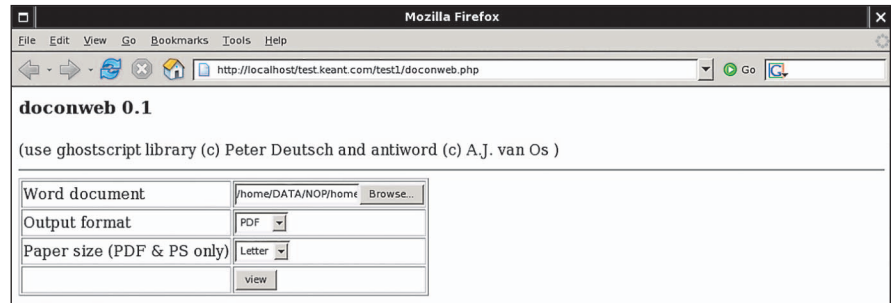


**GLX OpenShop™**  
**ITCenter Kuningan**  
 Lower Ground B1-21  
 Jl. Prof. Dr. Satrio  
 Jakarta 12940  
 Ph. 021-5793-4060  
 Fax 021-5793-5557  
[info@gudanglinux.net](mailto:info@gudanglinux.net)

```

document</td><td><input type=
'file' name='worddoc'></
td></tr>
<tr><td>Output format</td><td>
<select name='output'>
<option value='PDF'>
PDF</option>
<option value='PS'>
PS</option>
<option value='TEXT'>
TEXT</option>
</select>
</td></tr>
<tr><td>Paper size (PDF & PS
only)</td><td>
<select name='paper'>
<option value=
'letter'>Letter</
option>
<option value='a4'>
A4</option>
</select>
</td></tr>
<tr><td>&nbsp;</td><td><input
type='submit' value='view'></
td></tr>
</table>
</form>
";
echo "
</body>
</html>
";
}
else
{
if ($_FILES['worddoc']['error']
!= 0)
{
header("Location: {$_SERVER
['PHP_SELF']}");
}
else
{
$output_arg = "";
$header_type = "";
switch ($_POST['output'])
{
case "PDF":
$output_arg = " -m
mapping_file -p{$_POST
['paper']} {$_FILES['worddoc']

```



Aplikasi doconweb.

```

['tmp_name']] | ps2pdf - >
{$_FILES['worddoc']['tmp_
name']}.pdf && cat {$_FILES
['worddoc']['tmp_name']}.pdf";
$header_type =
"Content-type: application/pdf";
break;
case "PS":
$output_arg = " -m
mapping_file -p
{$_POST['paper']} {$_
FILES['worddoc']['tmp_
name']}";
$header_type =
"Content-type:
application/postscript";
break;
case "TEXT":
$output_arg = " {$_
FILES['worddoc']['tmp_
name']}";
$header_type =
"Content-type: text/
plain";
break;
}
header($header_type);
system("$converter_app
$output_arg");

if ($_POST['output'] ==
"PDF") unlink($_FILES
['worddoc']['tmp_name'] .
".pdf");
}
}
?>

```

**Penjelasan source code:**

- Pertama-tama, melalui variabel \$converter\_app, kita mengatur path ke lokasi

program yang kita gunakan, yaitu antiword. Di komputer penulis, antiword berada di /usr/bin. Sementara, variabel \$mapping\_file berisikan nama file yang digunakan untuk pemetaan karakter. Karena antiword mengharuskan nama file dan bukan path lengkap, maka kita harus menyebutkan nama filenya saja, yang tersimpan di direktori /usr/share/antiword/. Pemetaan karakter digunakan untuk output ke PS dan PDF.

- Setelah itu, kita memeriksa variabel \$\_POST. Apabila belum diset, maka form harus ditampilkan. Di form ini, kita menyediakan fasilitas upload kepada user, beserta parameter lain untuk konversi seperti tipe output dan ukuran kertas. Tipe encode form adalah multipart/form-data yang dapat digunakan untuk fasilitas upload.
- Apabila pada pemeriksaan form sebelumnya variabel \$\_POST telah diset, maka berarti user telah men-submit form tersebut. Kita pun memeriksa apakah terjadi error pada saat meng-upload. Error bisa terjadi karena beberapa hal. Salah satunya adalah karena user tidak menyebutkan nama filenya. Atau, terjadi kegagalan transmisi. Informasi file-file yang diupload disimpan di dalam \$\_FILES[<NAMA VARIABEL FORM UNTUK TYPE=FILE>]. \$\_FILES merupakan associative array sehingga memungkinkan kita untuk memeriksa key error untuk \$\_FILES['worddoc']. Apabila bukan nol, maka terjadi kesalahan. Ketika terjadi kesalahan, kita mengirim header untuk membawa kita kembali ke halaman kita sendiri (yang karena \$\_POST belum diset, maka form kembali ditampilkan).
- Apabila file telah diupload dengan lancar, kita memeriksa tipe output yang diinginkan. Rencananya, kita akan meng-



gunakan cara kerja seperti berikut. Kita memiliki dua variabel: \$header\_type dan \$soutput\_arg. Variabel pertama berguna untuk melewati tipe dokumen (mime type) untuk header HTML, yang akan berujung bagi browser untuk menindaklanjuti sesuai tipe file-nya. Misal, tipe HTML akan ditampilkan sebagai HTML, tipe TEXT akan ditampilkan sebagai TEXT dan tipe lain sesuai pengaturan di browser. Tentunya, untuk ketiga output yang diinginkan, tipe dokumennya berbeda. Untuk TEXT, kita menggunakan text/plain, untuk PDF, kita menggunakan application/pdf dan untuk PS, kita menggunakan application/postscript. Sementara, variabel \$soutput\_arg berisikan semua arguman yang ingin dilewatkan ke program antiword. Dengan demikian, setelah kita mendapatkan isi kedua variabel tersebut pada blok switch(), kita pun tinggal memanggil fungsi header() untuk melewati tipe header dan fungsi system() untuk menjalankan program antiword beserta semua parameternya.

Khusus untuk output berupa format PDF, kita menghapus file sementara ketika perintah system() selesai dikerjakan.

- Perhatikanlah isi dari \$soutput\_arg. Setiap file yang di-upload oleh user akan diberikan nama *random* tertentu. Nama tersebutlah yang disimpan sebagai nama file untuk file yang di-upload ke harddisk server. Key untuk array \$\_FILES['worddoc'] adalah tmp\_name. Untuk format output berupa TEXT, maka \$soutput\_arg hanya akan berisi nama file random tersebut. Untuk format output berupa PS, maka akan berisi -m \$mapping\_file -p {\$POST['paper']} dan nama file random tersebut. Untuk format output berupa PDF, kita melewati output PS ke program ps2pdf dan menyimpan ke nama file random ditambah ekstensi pdf dan setelah itu memanggil program cat untuk menampilkan isi file PDF ke standard output, yang akan ditangkap oleh browser.
- Ada beberapa hal yang mungkin sedikit mengganggu. Pertama, tentunya harus diperhatikan pula maksimal ukuran file

yang diperbolehkan untuk di-upload (Anda bisa membaca manual PHP untuk lebih lengkapnya). Kedua, dengan code seperti ini, orang-orang bisa saja membuat robot untuk memanggil aplikasi kita yang akan berujung pada kegagalan service (DoS) karena server kelebihan tugas. Ada beberapa cara yang bisa dilakukan, misalnya dengan meminta user untuk memasukkan teks pada gambar yang selalu digenerate secara random (Anda bisa membaca resep pemrograman PHP milik Steven Haryanto untuk lebih jelasnya). Secara umum, program ini masih memerlukan banyak perbaikan untuk diterapkan pada jaringan publik.

Bisa kita lihat, dengan memanfaatkan tool-tool open source, kita bisa membangun solusi yang mungkin akan cukup mahal (bahasa pemrograman, web server, konverter, pembuat PDF umumnya bukan barang murah di dunia *proprietary*) apabila kita menggunakan sistem *proprietary*. ☺

**Noprianto** (noprianto@infolinux.co.id)

## Pelopor Training LINUX Indonesia Mewujudkan Kepercayaan Anda dengan:

● **Mencetak PROFESIONAL LINUX**  
Yang mampu membangun server dan mengelola jaringan komputer berbasis Linux & penerapan aplikasi pada internet/intranet beserta system security-nya.

# Program Intensif LINUX Profesional

Linux Server Development ⌚ 200 jam

### MATERI :

- **Hardware & Jaringan**  
Komponen & Konfigurasi Komputer, Setup & Membuat LAN
- **Linux Fundamental**  
Basic User, X-Window, System Administration & Networking
- **Internet & Aplikasi WEB**  
Browser, Search Engine, Email, FTP, HTML, CSS & JavaScript
- **Shell Programming**
- **Advanced System Administration**
- **Advanced Networking Administration**
- **PHP & MySQL**

### KARIER KELULUSAN:

- IT Division pada perusahaan yang menggunakan komputer dlm menyelesaikan pekerjaan sehari-hari
- Support System Administration & Networking
- ISP (Internet Service Provider) ● WEB Hosting
- Warant Development ● System Integration

### FASILITAS

- ☺ Ruang kuliah full AC
- ☺ Tiap peserta 1 PC, Min PIII, Ram 128, Network & Multimedia
- ☺ Internet
- ☺ Modul pelatihan
- ☺ Cotton bag
- ☺ T-Shirt exclusive
- ☺ CD Linux
- ☺ Block notes
- ☺ Disket
- ☺ Sertifikat

### JADUAL

- ☞ Senin s.d. Kamis (3 bulan)
- ☞ Sabtu & Minggu (6 bulan)
- ☞ Jam: 08.00 s.d 12.00 WIB.
- ☞ Jam: 13.30 s.d. 18.00 WIB.

### Migrasi Ke Linux

- Server
- Network & System Support
- Web & Desktop Application



LPKNF

## LEMBAGA PENDIDIKAN KOMPUTER NURUL FIKRI

- Jl. Margonda Raya No. 522 - Depok ☎/Fax. (021) 7874223-24, 77206991
- Jl. Mampang Prapatan X/4 Jakarta Selatan ☎ (021) 7947115, 7975235
- Jl. A. Yani - Sentra Niaga B.I/12, Bekasi ☎/Fax. (021) 88956879, 8853537

www.pilp.web.id  
www.nurulfikri.com  
info@nurulfikri.com

BERKUALITAS

TERPERCAYA  
PILP  
LPKNF