

# Hello World!

It's Time to Share

## My Pages

[Beranda](#)

[PHP](#)

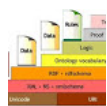
[Java & Android](#)

[Temu Kembali Informasi](#)

[About](#)

## Popular Articles

[Aplikasi Toko Buku Berbasis Java](#)



[Apa itu Ontology?](#)



[Peringkasan Teks Otomatis Berbahasa Indonesia : Part 1](#)



[Part 1 :: Membuat Search Engine Menggunakan Sphinx dan PHP](#)



[Part 2 :: Membuat Search Engine Menggunakan Sphinx dan PHP](#)

## My Words

► [2014](#) (5)

► [2013](#) (3)

▼ [2012](#) (9)

► [Desember](#) (4)

► [November](#) (1)

▼ [Oktober](#) (4)

[Part 4 :: Membuat Search Engine Menggunakan Sphinx...](#)

[Part 3 :: Membuat Search Engine Menggunakan Sphinx...](#)

[Part 2 :: Membuat Search Engine Menggunakan Sphinx...](#)

[Part 1 :: Membuat Search Engine Menggunakan Sphinx...](#)

## My Quote

It's fun to talk about dreams.  
All we can do is just believe and make

Sabtu, 20 Oktober 2012

## Part 1 :: Membuat Search Engine Menggunakan Sphinx dan PHP

Mencoba menghidupkan kembali blog ini, soalnya pernah buat blog beberapa kali karena lupa domainnya apa 6(-.-"). Lagian domain ini paling bagus namanya hehehe.



Sesuai dengan judulnya, kali ini saya berusaha membuat sebuah tutorial mengenai *search engine*. Umumnya teman-teman sudah cukup tahu apa yang dimaksud *search engine*, contohnya adalah google, yahoo, bing, dan lain-lain. Nah, kita akan mencoba membuat *search engine* milik kita sendiri dengan konten yang juga kita tentukan sendiri. Proses pembuatannya akan menggunakan bahasa pemrograman PHP dan sebuah *tools* yang bernama Sphinx. Tutorial ini dibagi menjadi beberapa bagian agar tidak terlalu panjang dan dapat dipahami (jadi teman-teman bisa mengembangkan lebih baik).

Sphinx merupakan *tools* yang lumayan lengkap untuk menciptakan sebuah *search engine*. Sphinx menyediakan berbagai pengaturan dan program untuk melakukan proses *indexing* (nanti akan dijelaskan) terhadap konten *search engine* yang kita inginkan. Sphinx juga dapat digunakan untuk beberapa bahasa pemrograman, seperti Java, Ruby, Perl, dan gatau lagi.

Sebelum mulai, siapin alat dan bahan yang diperlukan, [download XAMPP](#) (kalo belum punya) versi berapa aja. Kemudian [download Sphinx](#) versi terbaru aja (pilih sesuai OS Windows-nya). XAMPP untuk pembuatan web menggunakan PHP dan Sphinx untuk pembuatan *search engine*. Oh iya ini tutorialnya untuk Windows yak, kalau OS lain beda lagi struktur foldernya.

Lakukan instalasi untuk XAMPP sampai selesai (konfigurasi *default* aja), sedangkan untuk Sphinx cukup ekstraksi aja ke tempat yang mudah dijangkau, misal di partisi C:\. Kalau saya, saya taruh dalam folder baru bernama "sphinx", sehingga menjadi C:\sphinx. Nah, sekarang udah siap dimulai nih. #tepoktangan

Langkah awal yang dilakukan adalah menyiapkan **korpus** (yang akan menjadi ruang lingkup pencarian *search engine*) dalam bentuk *file* format .xml, sebenarnya bisa juga diambil dari *database* tetapi sekarang yang simpel dulu (maklum *newbie*). Nah bagaimana bentuk *file* xml-nya? Kurang lebih seperti ini :

```
1 <sphinx:docset>
2 <sphinx:schema>
3 <!-- Perhatikan! Tag sphinx:field ini menunjukkan tag-tag yang akan di indeks oleh sphinx -->
4 <sphinx:field attr="string" name="nama-tanaman"></sphinx:field>
5 <sphinx:field attr="string" name="nama-latin"></sphinx:field>
6 <sphinx:field attr="string" name="uraian"></sphinx:field>
7 <sphinx:field attr="string" name="nama-lokal"></sphinx:field>
8 <sphinx:field attr="string" name="khasiat"></sphinx:field>
9 </sphinx:schema>
10
11 <!-- Perhatikan! sphinx:document menunjukkan nomor dokumen artinya konten Asam Jawa merup -->
12 <sphinx:document id="1">
13 <nama-tanaman>Asam Jawa</nama-tanaman>
14 <nama-latin>Tamarindus indica, Linn.</nama-latin>
15 <uraian>Asam jawa (tamarindus indica) merupakan sebuah kultivar daerah tropis dan termas
16 <nama-lokal>Tamarind (Inggris), Tamarinier (Perancis), Asam Jawa (Indonesia), Celangi, T
17 <khasiat>Asma, Batuk, Demam, Sakit panas, Reumatik, Sakit perut, morbili, Alergi/biduren
18 </sphinx:document>
19
20 <sphinx:document id="2">
21 <nama-tanaman>Jambu Biji</nama-tanaman>
22 <nama-latin>Psidium guajava, Linn.</nama-latin>
23 <uraian>Jambu Biji (Psidium guajava) tersebar meluas sampai ke Asia Tenggara termasuk In
24 <nama-lokal>Psidium guajava (Inggris/Belanda), Jambu Biji (Indonesia), Jambu klutuk, Bay
25 <khasiat>Diabetes melitus, Maag, Diare (sakit perut), Masuk angin, Beser, Prolapsisani,
26 </sphinx:document>
27
28 <sphinx:document id="3">
29 <nama-tanaman>Jahe</nama-tanaman>
30 <nama-latin>Zingiber officinale Rosc.</nama-latin>
31 <uraian>Tanaman herba semusim, tegak, tinggi 40-50 cm. Batang semu, beralur, membentuk r
32 <nama-lokal>Zingiberis Rhizoma, Rimpang Jahe.</nama-lokal>
33 <khasiat>Asi, Batuk, Membangkitkan nafsu makan, Mulas, Perut kembung, Serbat, Gatal (ob
34 </sphinx:document>
35
36 <sphinx:document id="4">
37 <nama-tanaman>Lidah Buaya</nama-tanaman>
38 <nama-latin>Aloe Vera Linn.</nama-latin>
```

it happen.

#### My Team

My Team



#### My Campus



#### Me

Andre Fadila

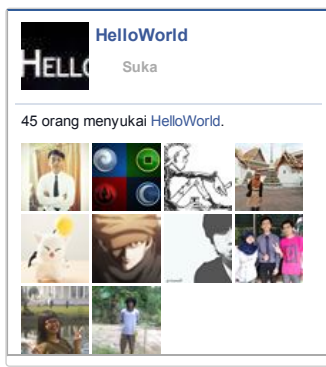
google.com/+AndreFadila



Ikuti

256 pengikut

#### My Friends

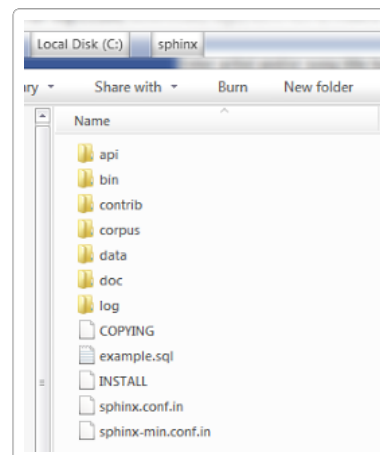


#### My Friends

```
39 <uraian>Tumbuhan liar di tempat yang berhawa panas atau ditanam orang di pot dan pekaran
40 <nama-lokal>Lidah buaya (Indonesia), Crocodiles tongues (Inggris), Jadam (Malaysia), Sal
41 <kehasiat>Shampo, minuman, Obat cacing, Luka bakar, Bisul, Luka bernanah, Amandel, Sakit
42 </sphinx:document>
43
44 <sphinx:document id="5">
45 <nama-tanaman>Mahkota Dewa</nama-tanaman>
46 <nama-latin>Phaleria macrocarpa [Scheff.] Boerl.</nama-latin>
47 <uraian>Mahkota dewa bisa ditemukan ditanam di pekarangan sebagai tanaman hias atau di k
48 <nama-lokal>Simalakama (Melayu), makutadewa, makuto mewa, makuto ratu, makuto rojo (Jawa
49 <kehasiat>disentri, psoriasis, jerawat, penyakit kulit, seperti ekzim dan gatal-gatal</kh
50 </sphinx:document>
51
52 <sphinx:document id="6">
53 <nama-tanaman>Kumis Kucing</nama-tanaman>
54 <nama-latin>Orthosiphon aristatus (Bl) Miq.</nama-latin>
55 <uraian>Terna, tumbuh tegak, pada bagian bawah berakar di bagian buku-bukunya, tinggi 1-
56 <nama-lokal>Mamang besar (Indonesia), Kutun, mamam, bunga laba-laba (Jawa), Mao Xu Cao (
57 <kehasiat>Infeksi Ginjal, Infeksi Kandung kemih, Kencing batu, Encok, Peluruh air seni, m
58 </sphinx:document>
59
60 <sphinx:document id="7">
61 <nama-tanaman>Kencur</nama-tanaman>
62 <nama-latin>Kaempferia galanga, Linn.</nama-latin>
63 <uraian>Kencur (Kaempferia galanga) termasuk suku tumbuhan Zingiberaceae dan digolongkan
64 <nama-lokal>Kencur (Indonesia, Jawa), Cikur (Sunda), Ceuko (Aceh), Kencor (Madura), Ceku
65 <kehasiat>Radang Lambung, Radang anak telinga, Influenza pada bayi, Masuk angin, Sakit Ke
66 </sphinx:document>
67
68 <sphinx:document id="8">
69 <nama-tanaman>Cengkeh</nama-tanaman>
70 <nama-latin>Syzygium aromaticum, (Linn.) Merr.</nama-latin>
71 <uraian>Cengkeh (Syzygium aromaticum) termasuk jenis tumbuhan perdu yang dapat memiliki
72 <nama-lokal>Clove (Inggris), Cengkeh (Indonesia, Jawa, Sunda), Wunga Lawang (Bali), Ca
73 <kehasiat>Kolera, Menghitamkan alis mata, Menambah denyut Jantung, Campak</kehasiat>
74 </sphinx:document>
75
76 <sphinx:document id="9">
77 <nama-tanaman>Kayu manis</nama-tanaman>
78 <nama-latin>Cinnamomum burmani (nees) Bl.</nama-latin>
79 <uraian>Pohon tinggi dapat mencapai 15 meter. Batang berkayu dan bercabang-cabang. Daun
80 <nama-lokal>Cinnamomi Cortex, Cassia vera, Kayu Manis (Padang), Kayu Manis, cinnamon (in
81 <kehasiat>Mencret, Membangkitkan nafsu makan, Memberi aroma pada makanan dan obat tradisi
82 </sphinx:document>
83
84 <sphinx:document id="10">
85 <nama-tanaman>Bunga Matahari</nama-tanaman>
86 <nama-latin>Helianthus annuus Linn.</nama-latin>
87 <uraian>Herba anual (umumnya pendek, kurang dari setahun), tegak, berbulu, tinggi 1 - 3 m
88 <nama-lokal>bungngong matahari, bungka matahari, purbanegara, Bunga panca matoari, bung
89 <kehasiat>Hipertensi, Sakit kepala, Sakit gigi, nyeri menstruasi, reumatik, Nyeri lambung
90 </sphinx:document>
91
92 <!-- Nah kalau mau nambah lagi dokumennya atau mau beda kontennya tinggal ganti aja isi f
93 </sphinx:docset>
```

Korpus ini berisi dokumen tanaman obat, kemudian simpan dengan nama file "korpus.xml". Kita taruh file korpus ini ke dalam folder "sphinx", saya buat folder baru bernama "corpus" dan pindahkan file xml tadi ke dalamnya sehingga strukturnya menjadi "C:\sphinx\corpus\korpus.xml".

Sekarang kita masuk ke proses *indexing*, yang harus dilakukan pertama kali adalah mengatur konfigurasi *indexing* Sphinx. Sebelumnya kita siapkan dua buah folder baru lagi, kita beri nama "data" dan "log" untuk keperluan *indexing*. Sehingga struktur folder yang kita miliki sekarang kurang lebih seperti ini.



Saatnya kita melakukan konfigurasi Sphinx, kita akan mengubah isi file **sphinx-min.conf.in** (konfigurasi minimal). Hal ini sebenarnya hanya menambahkan sebuah fungsi dan parameter yang ada sebelumnya :

```
1 #
2 # Minimal Sphinx configuration sample (clean, simple, functional)
```

?

Andre Fadila

+ ke lingkaran



256 memiliki saya di  
lingkaran

Lihat  
semua

```
3 #
4
5 source src1
6 {
7   type = mysql
8   sql_host = localhost
9   sql_user = test
10  sql_pass =
11  sql_db = test
12  sql_port = 3306 # optional, default is 3306
13  sql_query = \
14    SELECT id, group_id, UNIX_TIMESTAMP(date_added) AS date_added, title, content \
15    FROM documents
16  sql_attr_uint = group_id
17  sql_attr_timestamp = date_added
18  sql_query_info = SELECT * FROM documents WHERE id=$id
19 }
20
21 # Fungsi untuk korpus bentuk file xml
22 source srcxml
23 {
24   type = xmlpipe
25   xmlpipe_command = type C:\sphinx\corpus\korpus.xml # Letak file korpus
26   xmlpipe_field = docno
27   xmlpipe_field = title
28   xmlpipe_field = content
29   xmlpipe_fixup_utf8 = 1
30 }
31
32 # Fungsi indexing
33 index test1
34 {
35   source = srcxml # Memilih fungsi korpus yang akan digunakan
36   path = C:/sphinx/data/test1
37   docinfo = extern
38   #charset_type = sbcs
39   charset_type = utf-8
40
41   min_word_len = 3 # Minimum panjang kata yang akan diindeks
42   enable_star = 0
43   html_strip = 0
44   stopwords = C:/sphinx/data/stopwords.txt # Daftar kata buang
45   min_stemming_len= 4 # Minimum panjang kata yang akan dipotong
46 }
47
48
49 index testrt
50 {
51   type = rt
52   rt_mem_limit = 32M
53   path = C:/sphinx/data/testrt
54   charset_type = utf-8
55   rt_field = title
56   rt_field = content
57   rt_attr_uint = gid
58 }
59
60
61 indexer
62 {
63   mem_limit = 32M
64 }
65
66
67 searchd
68 {
69   listen = 9312
70   # listen = 9306:mysql41
71   log = C:/sphinx/log/searchd.log
72   query_log = C:/sphinx/log/query.log
73   read_timeout = 5
74   max_children = 30
75   pid_file = C:/sphinx/log/searchd.pid
76   max_matches = 1000
77   seamless_rotate = 1
78   preopen_indexes = 1
79   unlink_old = 1
80   workers = threads # for RT to work
81   binlog_path = C:/sphinx/data
82 }
```

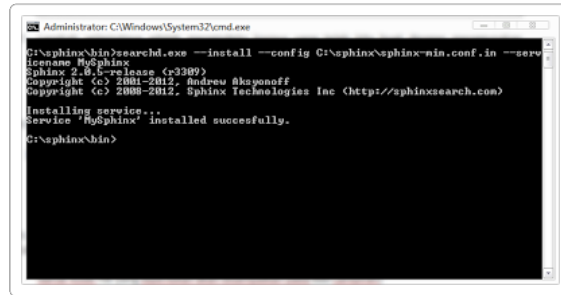
Kemudian barulah kita akan menjalankan program *indexing* Sphinx yang bernama **indexer.exe** menggunakan *command prompt* (cmd) dengan hak akses Administrator (*Run as Administrator*). Berikut panduannya kalau belum memahami *command prompt* :

```
Administrator: C:\Windows\System32\cmd.exe
Microsoft Windows [Version 6.1.7600]
Copyright (c) 2009 Microsoft Corporation. All rights reserved.

C:\Windows>cd..
C:\>cd sphinx\bin
C:\sphinx\bin>indexer.exe --config C:\sphinx\sphinx-mn.conf.in --all
Sphinx 2.0.6-release (c3380)
Copyright (c) 2001-2012, Andrew Knyazov
Copyright (c) 2008-2012, Sphinx Technologies Inc (http://sphinxsearch.com)
Using config file 'C:\sphinx\sphinx-mn.conf.in'...
Indexing index 'test1':
WARNING: source 'srcxml': both embedded and configured schemas found; using embed
ded (client), pos=1, docid=0
Collected 10 docs, 0.0 MB
Sorted 0.0 MB; 100.0% done
Total 10 docs, 12954 bytes
Total 0.010 sec, 118194 bytes/sec, 912.40 docs/sec
Skipping non-plain index 'testrt'...
Total 2 reads, 0.000 sec, 5.4 kb/call avg, 0.0 nsec/call avg
Total 9 writes, 0.000 sec, 4.0 kb/call avg, 0.0 nsec/call avg
C:\sphinx\bin>
```

Hal yang pertama dilakukan adalah pindah posisi folder menuju posisi folder **sphinx** berada. Kemudian eksekusi **indexer.exe** dengan parameter konfigurasi **--config** menggunakan file **sphinx-min.conf.in** yang barusan kita *edit*. Jika tidak ada pesan **ERROR** artinya proses *indexing* selesai dan benar.

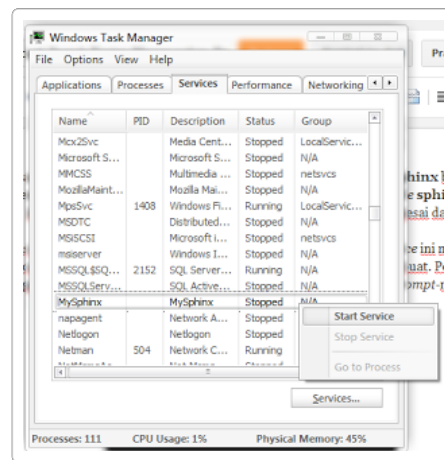
Setelah proses indeks, kita akan membuat *service* Windows untuk Sphinx. *Service* ini nantinya digunakan sebagai pemroses *search engine* yang memanfaatkan indeks yang barusan kita buat. Pembuatan *service* ini juga menggunakan *command prompt*, jadi jangan ditutup dulu *command prompt*-nya. Berikut perintahnya :



```
Administrator: C:\Windows\System32\cmd.exe
C:\sphinx\bin>searchd.exe --install --config C:\sphinx\sphinx-min.conf.in --servicename MySphinx
Sphinx 2.0.8-release (r3389)
Copyright (c) 2001-2012, Andrew Askayonoff
Copyright (c) 2008-2012, Sphinx Technologies Inc (http://sphinxsearch.com)
Installing service...
Service 'MySphinx' installed successfully.
C:\sphinx\bin>
```

Dapat dilihat, untuk membuat *service* Sphinx menggunakan program **searchd.exe** dengan parameter konfigurasi **sphinx-min.conf.in**, *service* yang barusan dibuat diberi nama **MySphinx** (perhatikan parameter **--servicename**). Jika tidak ada pesan **ERROR** artinya proses pembuatan *service* telah berhasil.

Langkah terakhir Part 1 ini adalah menjalankan *service* yang barusan kita buat. Cara yang paling mudah menjalankannya lewat **Task Manager**, buka tab **Services** cari **MySphinx** klik kanan pilih **Start Service**.



Ketika statusnya sudah berubah **Running** artinya *service MySphinx* sudah berjalan. Oke sudah selesai Part 1, kita lanjut ke **Part 2**, waktunya membuat web untuk *search engine* kita. :)


 +3 Rekomendasikan ini di Google

Diposkan oleh Andre Fadila

Label: **information retrieval**, **php**, **search engine**, **sphinx**, **temu kembali informasi**, **website**

Reaksi: ga penting (0) lumayan (0) tambah lagi (1)

## 7 komentar:

 Allyn 21 Juni 2013 03.54

Trimakasih, posting-an ini sangat membantu saya dalam mengerjakan tugas Sistem Temu Balik Informasi..

Balas

Balasan

 Andre Fadila 24 Juni 2013 20.27

Sama-sama pantau terus yak :D

 Dani Jazmi 10 Juli 2013 01.09

Gan . . .  
ane pas indexing korpusnya kok error ya ??  
ERROR: index 'test1': xmlpipe: expected ", got '

ada solusi ga ??  
untuk refrensi TA saya . . .



**Andre Fadila** 16 Juli 2013 05.27

Belum pernah dapet error kaya gitu gan, mungkin ada sesuatu di file XML nya.  
Kalau nyoba contoh yg di blog ini bisa gan?

[Balas](#)



**hamid21near** 10 Juni 2014 15.10

sama nih gan errornya,,error index test 1 jg  
agan pake sphinx dan xampp versi brp?

[Balas](#)

[Balasan](#)



**Muhammad Teguh** 23 Juni 2014 20.57

*Komentar ini telah dihapus oleh pengarang.*

[Balas](#)



**Omen Dyn** 5 November 2014 01.29

Kueri gagal : connection to localhost:9312 failed (errno=10061, msg=No connection could be made because the target machine actively refused it.).

[Balas](#)

Masukkan komentar Anda...

Beri komentar sebagai: Google Accour ▼

[Publikasikan](#)

[Pratinjau](#)

[Posting Lebih Baru](#)

[Beranda](#)

Langganan: [Poskan Komentar \(Atom\)](#)