

# SOTAS : APLIKASI PENGINDEKS PDF TUGAS AKHIR DAN SKIRPSI BERBASIS SOLR-PHP

**Ahmad Fajar Arafat**

D3 Manajemen Informatika, Fakultas Teknik, Universitas Negeri Surabaya, ahmadarafat@mhs.unesa.ac.id

**I Kadek Dwi Nuryana**

Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya, dwinuryana@unesa.ac.id

## Abstrak

Dalam proses pengerjaan tugas akhir dan skripsi terkadang mahasiswa akan merasa bingung dalam mengambil judul yang akan di bahas. Dikarenakan kurangnya referensi judul yang akan atau telah dikerjakan sebelumnya. Saat ini ruang baca Jurusan Teknik Informatika Fakultas Teknik Universitas Negeri Surabaya memiliki banyak koleksi Tugas Akhir maupun Skripsi. Dengan banyaknya koleksi ini mahasiswa maupun dosen kesulitan dalam melakukan proses pencarian. Sehingga untuk mempermudah maka pada penelitian ini dibuatlah aplikasi layanan pengindeks file pdf Tugas Akhir dan Skripsi menggunakan Apache SOLR. Aplikasi SOTAS menggunakan Apache SOLR dalam pengindekss dan PHP sebagai aplikasi tampilan untuk user, aplikasi ini diharapkan untuk mempermudah baik dosen maupun mahasiswa dalam proses pencarian judul Tugas Akhir maupun Skripsi.

**Kata Kunci:** Apache Solr, Indeks, PDF, PHP

## Abstract

In the process of working on the final project and thesis, sometimes students will feel confused in taking the title that will be discussed. Due to the lack of references to titles that will be or have been worked on before. Currently, the reading room of the Department of Informatics, Faculty of Engineering, State University of Surabaya has a large collection of Final Projects and Theses. With this large collection, both students and lecturers have difficulty in carrying out the search process. So that to make it easier, in this study, a pdf file indexing service application was created using Apache SOLR. The SOTAS application uses Apache SOLR in indexing and PHP as a display application for users, this application is expected to make it easier for both lecturers and students in the process of searching for the title of a Final Project or Thesis.

**Keywords:** Apache Solr, Indeks, PDF, PHP

## PENDAHULUAN

Tugas akhir atau Skripsi adalah suatu karya ilmiah yang disusun oleh mahasiswa program sarjana berdasarkan hasil penelitian suatu masalah yang dilakukan secara seksama dengan bimbingan Dosen Pembimbing. Tugas Akhir atau Skripsi merupakan salah satu persyaratan yang harus dipenuhi untuk kelulusan program sarjana. Dalam mengerjakan Tugas Akhir atau Skripsi proses pertama adalah dengan pengajuan judul atau topik penelitian. Namun, pada porses ini terkadang mahasiswa akan merasa bingung dalam mengambil judul yang akan di bahas. Dikarenakan kurangnya referensi judul yang akan atau telah dikerjakan sebelumnya. Permasalahan acap kali terjadi di beberapa universitas dan perguruan tinggi yang belum memiliki perpustakaan fisik maupun digital. Saat ini di ruang baca Jurusan Teknik Informatika Fakultas Teknik Universitas Negeri Surabaya memiliki banyak koleksi Tugas Akhir maupun Skripsi. Dengan banyaknya koleksi ini mahasiswa maupun dosen kesulitan dalam melakukan proses pencarian. Sehingga untuk mempermudah maka pada penelitian ini dibuatlah aplikasi layanan pengindeks file pdf Tugas Akhir dan

Skripsi menggunakan Apache SOLR.

Berdasarkan permasalahan yang dijelaskan pada paragraf sebelumnya penulis bermaksud melakukan penelitian berjudul "SOTAS : Aplikasi pengindeks pdf Tugas Akhir dan Skripsi berbasis SOLR-PHP". Aplikasi SOTAS ini diharapkan untuk mempermudah baik dosen maupun mahasiswa dalam proses pencarian judul Tugas Akhir maupun Skripsi.

## KAJIAN PUSTAKA

### Apache Solr

**Solr** (diucapkan "solar") adalah platform pencarian perusahaan open source yang populer, super cepat, yang ditulis dengan bahasa Java dan dibangun dari proyek Apache LuceneTM pada tahun 2010. Solr sangat dapat diandalkan, dapat diukur, dan toleran terhadap kesalahan, menyediakan pengindeksan terdistribusi, replikasi, dan permintaan beban-seimbang, kegagalan dan pemulihan otomatis, konfigurasi terpusat, dan banyak lagi. Solr mendukung fitur pencarian dan navigasi dari banyak situs internet terbesar di dunia. (Admin, 2017)

Solr adalah server pencarian perusahaan yang berdiri

sendiri dengan sebuah REST seperti API. Dapat meletakkan dokumen di dalamnya (disebut "pengindeksan") melalui JSON, XML, CSV atau biner melalui HTTP. Dapat menanyakannya melalui HTTP GET dan menerima JSON, XML, CSV atau hasil biner.

Fitur – fitur yang ada di Apache Solr :

- a) Kemampuan pencarian teks lengkap tingkat lanjut, didukung oleh Lucene™, Solr memungkinkan kemampuan pencocokan yang kuat termasuk frasa, wildcard, penggabungan, pengelompokan dan banyak lagi di semua jenis data.
- b) Pengoptimalan untuk traffic bervolume tinggi, Solr terbukti pada skala yang sangat besar di seluruh dunia.
- c) Open Interfaces berbasis standar – XML, JSON, dan HTTP, Solr menggunakan alat yang Anda gunakan untuk membuat aplikasi membangun sebuah snap.
- d) Antarmuka administrasi yang komprehensif, Solr dikirim dengan antarmuka pengguna administratif yang terintegrasi dan responsif untuk memudahkan Anda mengontrol instansi Solr Anda.
- e) Pemantauan mudah, Solr menerbitkan banyak data metrik melalui JMX.
- f) Berskala tinggi dan toleransi kesalahan, dibangun berdasarkan uji coba Apache Zookeeper, Solr membuatnya mudah untuk skala naik turun. Solr menjadi replikasi, distribusi, penyeimbangan kembali, dan toleransi kesalahan di luar kotak.
- g) Fleksibel dan dapat diadaptasi dengan konfigurasi yang mudah, Solr dirancang untuk menyesuaikan dengan kebutuhan Anda sambil menyederhanakan konfigurasi.
- h) Pengindeksan yang hampir Real-Time, Solr memanfaatkan kemampuan Lucene, Real-Time Indexing, untuk memastikan Anda melihat konten Anda ketika Anda ingin melihatnya.
- i) Arsitektur plugin yang extensible, Solr menerbitkan banyak titik ekstensi yang terdefinisi dengan baik yang memudahkan plugin baik indeks dan query time plugins. Karena ini adalah open source yang berlisensi Apache, maka dapat mengubah kode apa pun yang diinginkan.
- j) Skema ketika diinginkan, tanpa skema ketika tidak diinginkan, menggunakan mode pemindahan data berbasis Solr ketika memulai dan kemudian menguncinya ketika saatnya untuk produksi.
- k) Ekstensi kuat, Solr mengirim dengan plugin opsional untuk mengindeks konten kaya (misalnya PDF, Word), deteksi bahasa, pengelompokan hasil pencarian, dan lainnya.
- l) Pencarian dan pemfilteran terstruktur, mencari data Anda sesuai keinginan menggunakan sebuah algoritma besar dari array pencarian.
- m) Pencarian geospasial, mengaktifkan pencarian berbasis lokasi sederhana dengan dukungan bawaan Solr untuk pencarian spasial.
- n) Analisis teks lanjut yang dapat dikonfigurasi, Solr mengirim dukungan untuk sebagian besar bahasa yang digunakan secara luas di dunia (Inggris, Cina, Jepang, Jerman, Perancis, dan banyak lagi) dan banyak alat analisis lain yang dirancang untuk membuat pengindeksan dan querying konten sefleksibel mungkin.
- o) Sangat terkonfigurasi dan cache pengguna yang dapat diperluas, fine-grained controls pada built-in cache Solr memudahkan untuk mengoptimalkan kinerja.
- p) Optimalisasi kerja, Solr telah disetel untuk menangani situs-situs terbesar di dunia.
- q) Keamanan yang dibuat didalamnya, mengamankan Solr dengan SSL, autentikasi dan peran berdasarkan otorisasi dan pluggable.
- r) Opsi penyimpanan lanjutan, dengan membangun kemampuan penyimpanan lanjutan Lucene (codec, direktori, dan lainnya), Solr mempermudah penyediaan kebutuhan penyimpanan data agar sesuai dengan aplikasi.
- s) Logging yang bisa dimonitor, mudah mengakses file log Solr dari antarmuka admin.
- t) Saran Query, ejaan, dan lainnya, Solr mengirimkan dengan kapabilitas lanjutan untuk pelengkapan otomatis (pencarian orang yang mencari tahu), pemeriksaan ejaan, dan lainnya.
- u) Solr mendukung JSON, CSV, XML.
- v) Pemetaan dokumen, Solr disertakan dengan Apache Tika, sehingga mudah untuk mengindeks konten seperti Adobe PDF, Microsoft Word, dan banyak lagi.
- w) Apache UIMA, Solr terintegrasi ke Apache UIMA, membuatnya mudah untuk memanfaatkan NLP dan alat-alat lain sebagai bagian dari aplikasi.
- x) Beberapa pencarian terindikasi, Solr mendukung arsitektur multi-tenant, sehingga mudah untuk mengisolasi pengguna dan konten.
- y) Apache Solr adalah sebuah alat hebat dengan kemampuan pencarian yang luar biasa. Untuk mencari dokumen, ia melakukan operasi berikut secara berurutan:
  1. Pengindeksan: Pertama, mengkonversi dokumen ke dalam format yang dapat dibaca oleh mesin yang disebut Indexing.
  2. Querying: Memahami persyaratan permintaan yang diminta oleh pengguna. Istilah-istilah ini bisa berupa gambar, kata kunci, dan banyak lagi.
  3. Pemetaan: Solr memetakan query pengguna ke dokumen yang disimpan dalam basis data untuk

menemukan hasil yang sesuai.

4. Peringkat hasilnya: Setelah mesin mencari dokumen yang diindeks, maka akan memberi peringkat output sesuai relevansinya.

### Aplikasi

Aplikasi berasal dari kata application yang artinya penerapan, lamaran, penggunaan. Secara etimologi, aplikasi merupakan program siap pakai yang dibuat untuk melaksanakan suatu fungsi bagi pengguna atau aplikasi yang lain dan dapat digunakan oleh sasaran yang dituju (Susanto, 2016). Aplikasi dapat dikatakan sebagai suatu perangkat lunak atau software yang dapat bekerja dengan menjalankan instruksi-instruksi dari pengguna (user), aplikasi banyak diciptakan untuk membantu berbagai aktivitas sehari-hari, contohnya seperti membuat laporan, percetakan, desain grafis, editing video, dan berbagai aktivitas lainnya.

### JSON

JSON (JavaScript Object Notation) adalah format pertukaran data yang ringan, mudah dibaca dan ditulis oleh manusia, serta mudah diterjemahkan dan dibuat (generate) oleh komputer. Format ini dibuat berdasarkan bagian dari Bahasa Pemrograman JavaScript, Standar ECMA-262 Edisi ke-3 - Desember 1999. JSON merupakan format teks yang tidak bergantung pada bahasa pemrograman apapun karena menggunakan gaya bahasa yang umum digunakan oleh programmer keluarga C termasuk C, C++, C#, Java, JavaScript, Perl, Python dll. Oleh karena sifat-sifat tersebut, menjadikan JSON ideal sebagai bahasa pertukaran-data. (Admin, 1999)

### XML

XML (Extensible Markup Language) adalah bahasa markup untuk keperluan umum yang disarankan oleh W3C untuk membuat dokumen markup keperluan pertukaran data antar sistem yang beraneka ragam. XML merupakan kelanjutan dari HTML (HyperText Markup Language) yang merupakan bahasastandar untuk melacak Internet. (Admin, 2019)

XML didesain untuk mampu menyimpan data secara ringkas dan mudah diatur. Kata kunci utama XML adalah data yang jika diolah bisa memberikan informasi. XML menyediakan suatu cara terstandarisasi namun bisa dimodifikasi untuk menggambarkan isi dari dokumen. Dengan sendirinya, XML dapat digunakan untuk menggambarkan sembarang viewdatabase, tetapi dengan satu cara yang standar.

### XAMPP

XAMPP adalah software aplikasi pengembang yang digunakan untuk pengembangan website berbasis PHP

dan juga sebagai server untuk lokal dalam pembuatan database dengan MySQL. XAMPP sendiri merupakan kepanjangan dari :

- a) X (OS X, maksudnya dapat dijalankan di berbagai sistem operasi / operating system / OS, seperti Windows, Linux, MacOS, dan Solaris).
- b) A (Apache, merupakan aplikasi web server. Tugas utama Apache adalah menghasilkan halaman web yang benar kepada user berdasarkan kode PHP yang dituliskan oleh pembuat halaman web. Jika diperlukan juga berdasarkan kode PHP yang dituliskan, maka dapat saja suatu database diakses terlebih dahulu (misalnya dalam MySQL) untuk mendukung halaman web yang dihasilkan).
- c) M (MySQL).
- d) P (PHP, merupakan Bahasa pemrograman web yang bersifat server-side-scripting).
- e) P (Perl, bahasa pemrograman untuk segala keperluan, dikembangkan pertama kali oleh Larry Wall di mesin Unix).

XAMPP berfungsi sebagai server lokal yang dapat berdiri sendiri (localhost), yang terdiri dari program Apache HTTP Server, MySQL database dan penerjemah bahasa yang ditulis dengan PHP dan Perl. Software atau aplikasi ini tersedia dalam GNU (General Public Lisence) dan bebas sehingga dapat membangun website dinamis dengan PHP secara offline karena XAMPP memiliki Apache sebagai server HTTP lokal. Selain itu juga dapat membuat database dengan adanya MySQL sebagai DBMS menggunakan phpMyAdmin, dimana merupakan software yang ditulis menggunakan Bahasa Pemrograman PHP untuk menangani administrasi dari MySQL melalui WWW (World Wide Web). (Tumanggor, 2019)

### METODE

#### SDLC Waterfall Model

Metode rekayasa yang diterapkan pada penelitian ini adalah SDLC (System Development Life Cycle) dengan menggunakan model waterfall. Tahapan dalam model ini meliputi analysis system requirement (analisis kebutuhan sistem), system design (desain sistem), system implementation (implementasi sistem), system verification/testing (uji coba sistem, dan system maintenance (pemeliharaan sistem).



Gambar 1. SDLC Waterfall Model

## Analisis Kebutuhan Sistem

### a) Pengumpulan Data

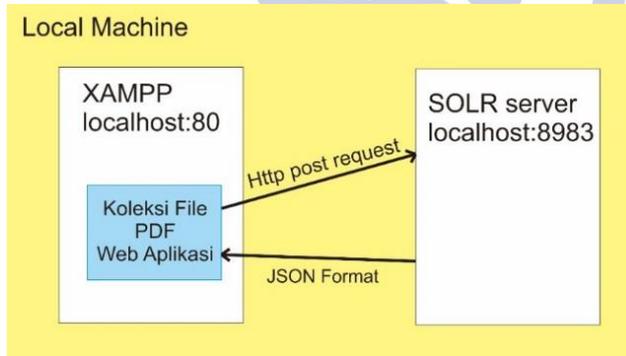
Penelitian ilmiah harus dilakukan teknik penyusunan yang sistematis untuk memudahkan langkah-langkah yang akan diambil. Penulis akan menggunakan salah satu metodologi penelitian sebagai langkah pertama dalam pengumpulan data, yaitu studi literatur terhadap hal-hal yang relevan dengan penelitiannya baik dari buku, jurnal, maupun karya ilmiah lainnya. Data atau informasi yang didapat dari studi literatur tersebut akan digunakan sebagai acuan untuk melanjutkan penelitian ke tahap berikutnya.

### b) Analisis Sistem

Pembahasan pada bagian ini akan melakukan analisis terhadap indeks file PDF menggunakan Apache Solr, dengan memanfaatkan library dari Apache Tika sebagai extract handler agar file pdf bisa langsung terbaca oleh Apache Solr. Serta menggunakan library dari solarium untuk keperluan web application sehingga mudah untuk menghubungkan PHP dengan Solr server.

## Desain Sistem

Pada tahapan ini merupakan rancangan mekanisme proses indexing file PDF dan mekanisme pencarian hasil indexing yang terbentuk oleh Apache Solr. Sistem yang dibuat akan berjalan pada mesin local sehingga web server serta Solr server terinstall pada mesin yang sama. Berikut gambar dari arsitektur desain yang akan dibuat.



Gambar 2. Desain Sistem

Pada gambar diatas web aplikasi terinstall pada XAMPP local web server yang berjalan pada localhost:80. Sedangkan Solr server yang terinstall berjalan pada localhost:8983. Ketika melakukan indeks dan pencarian web aplikasi melakukan komunikasi dengan Solr server dengan cara mengirimkan http post request dan Solr akan mengembalikan request dalam bentuk format JSON.

## HASIL DAN PEMBAHASAN

### Implementasi Sistem

Agar SOLR dapat mengekstrak file pdf, digunakan fitur request handler yang terdapat pada Apache Tika.

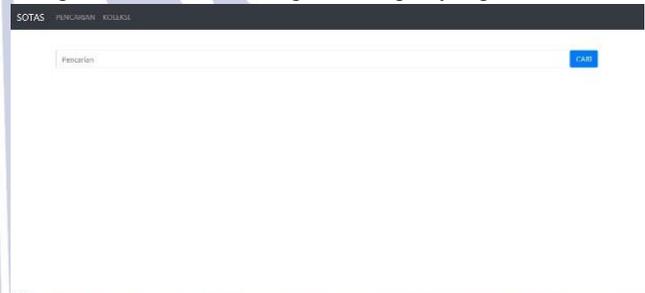
Dengan merubah konfigurasi pada file solrconfig.xml Berikut untuk konfigurasinya.

```
<requestHandler name="/update/extract" startup="lazy" class="solr.extraction.ExtractingRequestHandler" >
  <lst name="defaults">
    <str name="fmap.content">text</str>
    <str name="lowernames">true</str>
    <str name="uprefix">attr_</str>
    <str name="captureAttr">true</str>
    <str name="fmap.a">links</str>
    <str name="fmap.dlv">ignored</str>
  </lst>
</requestHandler>
```

Dan juga pada file schema menambahkan field yang dibutuhkan sebagai berikut.

```
<dynamicField name="attr_*"
  type="text_general" indexed="true"
  stored="true" multiValued="true"/>
```

Untuk Hasil tampilan beranda web aplikasi indeks sebagai berikut sesuai dengan rancangan yang dibuat.



Gambar 3. Tampilan Homepage

Pada server solr sudah terdapat beberapa dokumen yang sudah terindeks seperti gambar berikut.



Gambar 4. Server Solr

Untuk menghubungkan PHP dengan apache SOLR agar dapat mengakses data diperlukan konfigurasi seperti gambar berikut.

```
<?php
return [
  'endpoint' => [
    'localhost' => [
      'host' => env( key: 'SOLR_HOST', default: '127.0.0.1'),
      'port' => env( key: 'SOLR_PORT', default: '8983'),
      'path' => env( key: 'SOLR_PATH', default: '/'),
      'core' => env( key: 'SOLR_CORE', default: 'sotas')
    ]
  ]
];
```

Gambar 5. Connetion

Berikut adalah konfigurasi query update untuk mengupload file pdf ke dalam server solr sekaligus langsung mengekstrak file pdf agar server solr dapat membaca dan menyimpan.

```

$file_path = public_path( path: 'filepdf/APLIKASI_JARINGAN_SYARAF_TIRUAN.pdf');

$update = $this->client->createUpdate();
//create document
$doc = $update->createDocument();
$doc->id = 1;
$doc->some = 'more fields';

$extract = new Query();
$extract->addFieldMapping('content', 'text');
$extract->setUprefix( uprefic 'attr_');
$extract->setDefaultField('_text_');
$extract->setDocument($doc);
$extract->setFile($file_path);
$extract->setCommit( commit: true);
$extract->setOmitHeader( value: false);
$this->client->extract($extract);
$update->addDocument($doc);

//commit the update
$update->addCommit();

```

Gambar 6. Query Update

Sedangkan untuk query select pada server solr sebagai berikut konfigurasi yang diperlukan.

```

public function search()
{
    // get a select query instance
    $query = $this->client->createQuery($this->client::QUERY_SELECT);

    // pagination example
    $resultsPerPage = 15;
    $currentPage = 1;

    // set the number of results to return
    $query->setRows($resultsPerPage);

    // set the 0-based result to start from, taking into account pagination
    $query->setStart( start: ($currentPage - 1) * $resultsPerPage);

    // this executes the query and returns the result
    $resultset = $this->client->execute($query);

    // display the total number of documents found by solr
    echo 'NumFound: ' . $resultset->getNumFound();

    // show documents using the resultset iterator
    foreach ($resultset as $document) {

```

Gambar 7. Query Select

Ini adalah lanjutan dari query select untuk menampilkan hasil seluruh indeks.

```

    // show documents using the resultset iterator
    foreach ($resultset as $document) {

        echo '<hr/><table>';

        // the documents are also iterable, to get all fields
        foreach ($document as $field => $value) {
            // this converts multi-valued fields to a comma-separated string
            if (is_array($value)) {
                $value = implode( glue: ', ', $value);
            }

            echo '<tr><th> ' . $field . '</th><td> ' . $value . '</td></tr>';
        }

        echo '</table>';
    }
}

```

Gambar 8. Query Select

## PENUTUP

### Simpulan

Berdasarkan hasil ujicoba sistem yang telah dilakukan sebelumnya, penelitian yang berjudul *SOTAS : Aplikasi Pengindek PDF tugas Akhir dan Skripsi Berbasis SOLR-PHP* ini dapat mengindeks file pdf ke Apache Solr melalui web php.

### Saran

Masih banyak pengembangan sistem yang diperlukan untuk aplikasi SOTAS ini. Untuk lebih bagus lagi agar dapat highlight keyword yang dimasukkan pada hasil pencarian sehingga user tidak kebingungan untuk melihat hasil yang dicari.

## DAFTAR PUSTAKA

- Admin, 1999. *JSON*. [Online]  
Available at: <http://www.json.org>  
[Diakses 10 December 2019].
- Admin, 2017. *The Apache Software Foundation*. [Online]  
Available at: <http://lucence.apache.org/solr/>  
[Diakses 10 March 2018].
- Admin, 2019. *Wikipedia*. [Online]  
Available at: <http://id.wikipedia.org/wiki/XML>  
[Diakses 10 December 2019].
- Tumanggor, R., 2019. *Fungsi XAMPP dan Cara Menginstalnya*. [Online]  
Available at:  
<http://www.tifkom.net/2015/08/fungsi-xampp-dan-cara-menginstalnya.html>  
[Diakses 10 December 2019].



**UNESA**

Universitas Negeri Surabaya