**Sastrawi**

Sastrawi is a simple PHP library which allows you to reduce inflected words in Indonesian Language (Bahasa Indonesia) to their base form ([stem](http://en.wikipedia.org/wiki/Stemming)). Despite its simplicity, this library is designed to be high quality and well documented. For more information in english, see [README](https://github.com/andylib/sastrawi/blob/master/README.en.md).

| **Development** | **Master** | **Releases** | **Statistics** |
| --- | --- | --- | --- |
|  |  |  |  |

**Stemming**

[*Stemming*](http://en.wikipedia.org/wiki/Stemming) adalah proses mengubah kata berimbuhan menjadi kata dasar. Contohnya:

* menahan => tahan
* berbalas-balasan => balas

**Contoh kasus**

Katakanlah sebuah *blog post* berisi:

Rakyat memenuhi halaman gedung untuk menyuarakan isi hatinya.

Pencarian dengan *query* di bawah ini tidak akan menemukan *post* di atas,

SELECT \* FROM posts WHERE content LIKE '%suara%'

Proses *stemming* dapat membantu menemukan dokumen yang sedang dicari yaitu dengan menanggalkan imbuhan-imbuhan hingga hanya menyisakan kata dasar seperti berikut:

rakyat penuh halaman gedung suara isi hati

Lalu kata kunci pencarian juga dijadikan kata dasar:

Bersuara => suara

**Sastrawi**

* *Library PHP* untuk *stemming* Bahasa Indonesia.
* Mudah diintegrasikan dengan *framework* / *package* lainnya.
* Mempunyai *API* yang sederhana dan mudah digunakan.

**Demo**

<http://sastrawi.github.io>

**Cara Install**

Sastrawi dapat diinstall dengan [Composer](https://getcomposer.org).

1. Buka terminal (command line) dan arahkan ke directory project Anda.
2. [Download Composer](https://getcomposer.org/download/) sehingga file composer.phar berada di directory tersebut.
3. Tambahkan sastrawi ke file composer.json Anda :

php composer.phar require sastrawi/sastrawi:^1

Jika Anda masih belum memahami bagaimana cara menggunakan Composer, silahkan baca [Getting Started with Composer](https://getcomposer.org/doc/00-intro.md).

**Cara menjalankan test**

vendor/bin/phpunit

**Penggunaan**

Copy kode berikut di directory project anda. Lalu jalankan file tersebut.

<?php

// demo.php

// include composer autoloader

require\_once \_\_DIR\_\_ . '/vendor/autoload.php';

// create stemmer

// cukup dijalankan sekali saja, biasanya didaftarkan di service container

$stemmerFactory = new \Sastrawi\Stemmer\StemmerFactory();

$stemmer = $stemmerFactory->createStemmer();

// stem

$sentence = 'Perekonomian Indonesia sedang dalam pertumbuhan yang membanggakan';

$output = $stemmer->stem($sentence);

echo $output . "\n";

// ekonomi indonesia sedang dalam tumbuh yang bangga

echo $stemmer->stem('Mereka meniru-nirukannya') . "\n";

// mereka tiru

Menambah dan Mengurangi Kata Dasar

<?php

// include composer autoloader

require\_once \_\_DIR\_\_ . '/vendor/autoload.php';

// create stemmer

// cukup dijalankan sekali saja, biasanya didaftarkan di service container

$stemmerFactory = new \Sastrawi\Stemmer\StemmerFactory();

$dictionary = $stemmerFactory->createDefaultDictionary();

$dictionary->addWordsFromTextFile(\_\_DIR\_\_.'/my-dictionary.txt');

$dictionary->add('internet');

$dictionary->remove('desa');

$stemmer = new \Sastrawi\Stemmer\Stemmer($dictionary);

var\_dump($stemmer->stem('internetan')); //internet

**Pustaka**

**Algoritma**

Algoritma yang digunakan pada *library* ini adalah hak intelektual masing-masing pemiliknya yang tertera di bawah ini. Lalu untuk meningkatkan kualitas kode, algoritma tersebut diterapkan ke dalam *Object Oriented Design*.

* Algoritma Nazief dan Adriani
* Asian J. 2007. ***Effective Techniques for Indonesian Text Retrieval***. PhD thesis School of Computer Science and Information Technology RMIT University Australia
* Arifin, A.Z., I.P.A.K. Mahendra dan H.T. Ciptaningtyas. 2009. ***Enhanced Confix Stripping Stemmer and Ants Algorithm for Classifying News Document in Indonesian Language***, Proceeding of International Conference on Information & Communication Technology and Systems (ICTS)
* A. D. Tahitoe, D. Purwitasari. 2010. ***Implementasi Modifikasi Enhanced Confix Stripping Stemmer Untuk Bahasa Indonesia dengan Metode Corpus Based Stemming***, Institut Teknologi Sepuluh Nopember (ITS) – Surabaya, 60111, Indonesia

**Kamus Kata Dasar**

Proses stemming oleh Sastrawi sangat bergantung pada kamus kata dasar. Sastrawi menggunakan kamus kata dasar dari [kateglo.com](http://kateglo.com) dengan sedikit perubahan.

**Lisensi**

Lisensi sastrawi adalah MIT License (MIT) sedangkan lisensi kamus kata dasar dari Kateglo adalah [CC-BY-NC-SA 3.0](http://creativecommons.org/licenses/by-nc-sa/3.0/). Untuk informasi lebih lengkap silahkan lihat [Lisensi Sastrawi](https://github.com/sastrawi/sastrawi/blob/master/LICENSE) dan [Lisensi isi Kateglo](https://github.com/ivanlanin/kateglo#lisensi-isi).

**Di Bahasa Pemrograman Lain**

* [JSastrawi](https://github.com/jsastrawi/jsastrawi) - Java
* [cSastrawi](https://github.com/mohangk/c_sastrawi) - C
* [PySastrawi](https://github.com/har07/PySastrawi) - Python
* [Go-Sastrawi](https://github.com/RadhiFadlillah/go-sastrawi) - Go
* [Sastrawi-Ruby](https://github.com/meisyal/sastrawi-ruby) - Ruby

**Informasi Lebih Lanjut**

* [API Documentation](http://sastrawi.github.io/sastrawi-api-doc/master/)
* [FAQ](https://github.com/sastrawi/sastrawi/wiki/FAQ)
* [Wiki](https://github.com/sastrawi/sastrawi/wiki)
* [Roadmap](https://github.com/sastrawi/sastrawi/issues/milestones)
* [Bug Report, Questions, Ideas](https://github.com/sastrawi/sastrawi/issues)