**Stemming Bahasa Indonesia**

Stemming adalah proses untuk mengubah kata berimbuhan, maupun kata hasil dari proses tokeninasi ke bentuk kata dasarnya.

Algortima stemming yang kami gunakan sebagian besar bersumber dari Library PHP Stemming bahasa Indonesia yang disediakan di <https://github.com/sastrawi/sastrawi>. Struktur Code Program yang digunakan dalam library tersebut banyak yang **kami terjemahkan langsung** ke Program Stemming yang kami buat ini.

Ada beberapa proses yang kami ubah secara dasar, antara lain Pembacaan Rule Ambiguitas yang kami terjemahkan langsung dari Algoritma Nazief dan Adriani, Implementasi hashmap untuk dictionary lookup yang digunakan serta perubahan struktur rule reader agar memudahkan membaca rule dari file.

Berikut adalah FlowChart Stemming secara umum dalam program yang kami buat. Berikut adalah bagaimana proses stemming kami bekerja:



Ada beberapa tahapan dasar dari program stemming yang ada dalam tugas ini:

1. Pemeriksaan untuk kata-kata plural.
2. Pemeriksaan kata diluar rule umum yang ada seperti pelajar dan belajar.
3. Pemeriksaan jika kata tersebut hanya terdiri kurang dari 4 huruf.
4. Perubahan struktur kata melalui penghapusan Prefix maupun Suffix dari kata yang diperiksa
5. Hingga pada tahapan pemeriksaan akhir, yaitu perubahan prefix dan memeriksa kata yang telah dihapus suffixnya terlebih dahulu.

Seluruh tahapan diatas dilengkapi dengan dictionary lookup untuk memastikan jika kata awal maupun kata perubahan telah tersedia dalam kamus kata dasar.

**Pemeriksaan kata-kata plural**



Tahapan ini memeriksa apabila kata awal yang dimasukkan adalah kata plural/jamak. Jika kata tersebut kata plural maka kata tersebut akan dipisahkan terlebih dahulu menjadi kata individual, melalui aturan dalam flowchart diatas.

**Remove Prefix**



Tahapan ini menghapus prefix dari kata berdasarkan rule prefix yang sudah disediakan. Tahapan ini di loop tiga kali, karena diasumsikan bahwa setiap kata berimbuhan maksimal memiliki 3 prefix. Setiap perubahan dari penghapusan prefix, dilakukan pengecekan dikamus kata dasar untuk memastikan penghapusan prefix tidak merubah dan menyebabkan kata menjadi tidak beraturan.

**Remove Suffix**



Tahapan ini untuk menghapus dan memeriksa akhiran dari kata yang dihapus. Ada tiga jenis Suffix yang dihapus melalui tahapan ini, Possesive Prounoun, Particle dan Derivitional Suffix. Ketiga jenis suffix tersebut dihapus secara bertahap sesuai flowchart diatas.

**Pemeriksaan Akhir**



Tahapan ini dilakukan setelah penghapusan prefix dan suffix ditahap sebelumnya. Tahapan ini melalui **reversed removal checker.** Dimana kata yang sudah dihapus suffix-nya akan diperiksa lagi dengan melalui penghapusan prefix seperti pada flowchart diatas. Tahapan ini untuk memastikan bahwa ketika penghapusan suffix dan prefix tidak ada kata yang terlewatkan akibat perubahan struktur ditahapan sebelumnya.