

## BAB 6

### SIMPULAN DAN SARAN

#### 6.1 Simpulan

Berdasarkan pada hasil analisis yang sudah dilakukan, dapat ditarik sejumlah kesimpulan:

1. Sistem deteksi kesamaan konteks berhasil dibangun.
2. Berdasarkan hasil evaluasi, relevansi cluster website dengan website yang menggunakan bahasa inggris dan mandarin berhasil karena kata yang menjadi relevansi lebih banyak yang representative.
3. Berdasarkan hasil evaluasi, relevansi cluster website dengan software berhasil karena kata yang menjadi relevansi lebih banyak yang representative.
4. Berdasarkan hasil evaluasi, *cosine similarity* yang diterapkan berhasil karena hasil dari *cosine similarity* sama dengan hasil judgement.

#### 6.2 Saran

Berdasarkan hasil dan simpulan yang sudah diperoleh, terdapat beberapa saran dalam penelitian ini:

1. Diperlukan deteksi bahasa untuk setiap website yang akan di clustering.
2. Untuk software-based web page, deskripsi software dapat diperoleh dari google search result sebagai tambahan dari Wikipedia.
3. Diperlukan ekstraksi fitur pada gambar.
4. *Treshold cosine similarity* harus dikurangkan sesuai dengan kebutuhan, karena ditemukan banyaknya website yang seharusnya memiliki konteks yang sama namun dianggap berbeda oleh *cosine similarity*.
5. Diperlukan tokenisasi yang lebih akurat terhadap bahasa mandarin. Misalnya jieba [14].