

Kita sudah sama-sama mempelajari konsep umum dari pemrosesan bahasa alami (*natural language processing*) dan juga beberapa kakas yang bermanfaat seperti *POS Tagger*, *Syntactic Parser*, dan *Named-Entity Recognizer*.

Pada Tugas kali ini, kita akan melakukan *fine-grained sentiment analysis* pada level **aspek** (dari sebuah produk). Anda dituntut untuk **berpikir** bagaimana mengembangkan sistem sederhana untuk melakukan **ekstraksi aspek** dan juga menentukan *sentiment orientation* terhadap aspek tersebut.

Definisi Tugas

Agar memberikan gambaran yang jelas, sistem yang akan kita bangun menerima input berupa **sebuah kalimat** yang dapat berasal dari *review forum* (misal, Amazon) maupun media sosial yang lainnya. Diberikan sebuah kalimat **S**, kita akan mencoba untuk melakukan ekstraksi $\{(\text{aspect}_1, \text{SO}_1), (\text{aspect}_2, \text{SO}_2), \dots, (\text{aspect}_n, \text{SO}_n)\}$. SO_i adalah orientasi sentimen (positif/negatif) dari aspek ke- i .

Contoh#1

diberikan masukan berikut:

Brightness could be better and the viewing angle is not great , but for the price I paid for a 24inch monitor , this can not be beat .

keluaran yang dihasilkan sistem adalah:

Brightness[-1], **viewing angle**[-1], **price**[+1]

Contoh#2

diberikan masukan berikut:

The problems came from my video card .

keluaran yang dihasilkan sistem adalah:

video card[-1]

Contoh#3

diberikan masukan berikut:

The display is large enough , and the text rendering is sharp enough , that you 'll forget entirely about workspace issues .

keluaran yang dihasilkan sistem adalah:

display[+1], text rendering[+1]

Pada ketiga contoh di atas, **[+1]** menandakan orientasi positif terhadap aspek tersebut, dan **[-1]** adalah sebaliknya.

Dataset

Kita akan menggunakan dataset berbahasa inggris yang mengandung kalimat-kalimat *review* tentang tiga produk berikut: **komputer**, **router**, dan *speaker*. Dataset ini dibuat oleh Qian Liu dan Bing Liu (2015):

Qian Liu, Zhiqiang Gao, Bing Liu, Yuanlin Zhang. Automated Rule Selection for Aspect Extraction. In Proceedings of IJCAI '15, 2015.

Setiap kalimat pada dataset sudah diberi dilabeli dengan aspek-aspek beserta orientasi sentimen yang bersesuaian. Berikut adalah contoh untuk sebuah kalimat di dataset tersebut.

color[+1], price[+1] ## It has good color and price was great .

Kalimat di atas mengandung dua buah aspek: **color** dan **price**. Orientasi sentimen terhadap aspek “**color**” adalah positif, dan orientasi sentimen terhadap aspek “**price**” juga positif. Simbol yang digunakan pada dataset tersebut dijelaskan sebagai berikut:

1. **xxxx[+|-1]**: xxxx adalah **aspek** atau fitur dari sebuah produk
 - a. **[+1]**: sentimen positif
 - b. **[-1]**: sentimen negatif
2. **##**: awal dari sebuah kalimat. Setiap baris pada file merepresentasikan sebuah kalimat.

Sebenarnya, di dalam dataset, ada dua jenis aspek: **aspek eksplisit** dan **aspek implisit**. Aspek implisit lebih sulit untuk dideteksi karena *term*/kata aspek tersebut secara kasat mata tidak ditemukan pada sebuah kalimat. Untuk mempermudah sesi

hands-on kita kali ini, kita hanya menggunakan kalimat-kalimat yang mengandung aspek eksplisit saja.

Berikut adalah beberapa contoh kalimat yang lain. Silakan Anda pelajari **pola-pola** linguistik yang ada.

The sound quality has been poor
has markedly better stereo separation
I usually avoid horns
Also has a nice guest feature
This monitor works great
Sound is crystal clear
Sound is also nice
Sound quality is great
bass quality is beyond good
Really good sound quality
A little bit of control over the mids would be nice
a wonderful little router
sharp picture
The battery has a great life

Usulkan Arsitektur sistem yang mampu menangani Task tersebut!