

Panduan *Annotator* NLI Bahasa Jawa

Panduan ini bertujuan sebagai *Quality Assurance* (QA) terhadap hasil terjemahan dari bahasa Indonesia ke bahasa Jawa.

I. Persyaratan Peserta *Annotator*

- a. Peserta harus memahami bahasa Indonesia secara baik dan benar.
- b. Peserta harus memahami bahasa Jawa secara baik dan benar.
- c. Peserta merupakan warga asli suku Jawa, dan dapat secara aktif berbahasa Jawa.

II. Evaluasi *Annotator*

- a. Setiap *set* data terdiri dari sebuah kalimat premis dalam bahasa Indonesia dan pasangan terjemahan dalam bahasa Jawa, serta kalimat hipotesis dalam bahasa Indonesia dan pasangan terjemahannya dalam bahasa Jawa.
- b. Setiap *annotator* akan diberikan set data yang sama.
- c. Setiap *annotator* melakukan evaluasi terhadap set data yang diberikan dengan cara:
 - i. Menilai kualitas terjemahan antara 1-5 (5 paling bagus).
 - ii. Mengecek hasil terjemahan bebas dari kesalahan tata bahasa, ejaan, maupun tanda baca. Apabila terdapat kesalahan, maka *annotator* menandai data tersebut dengan memberikan nilai 1 pada kualitas terjemahan.
 - iii. Perbaikan terhadap hasil terjemahan pada bagian II.c.iii dilakukan oleh *annotator* independen sesuai dengan aturan pada bagian V.
 - iv. Mengecek apabila hasil terjemahan mengubah hubungan semantik antar kalimat.

III. Hubungan semantik antar kalimat

Hubungan semantik antar pasangan kalimat premis dan hipotesis dapat dilihat berdasarkan deskripsi berikut:

- a. Sepasang kalimat premis-hipotesis dapat dikatakan ***entailment*** apabila dapat disimpulkan bahwa kalimat hipotesis **benar** berdasarkan informasi yang ada pada kalimat premis.
- b. Sepasang kalimat premis-hipotesis dapat memiliki label ***contradiction*** apabila dapat disimpulkan bahwa kalimat hipotesis **salah** berdasarkan informasi yang ada pada kalimat premis.
- c. Sepasang kalimat premis-hipotesis dapat dikatakan ***neutral*** apabila berdasarkan informasi yang ada pada kalimat premis, kebenaran kalimat hipotesis tidak dapat ditentukan, atau tidak ada cukup informasi.

IV. Pemilihan Label Emas

Label emas (*gold label*) merupakan label utama yang akan dipakai untuk menandai hubungan semantik antara sebuah kalimat premis dan pasangan kalimat hipotesisnya. Hasil dari pengecekan hubungan semantik dari setiap *annotator* pada bagian II.c.iii akan dicek dengan cara:

- a. Apabila sepasang kalimat premis-hipotesis memiliki label yang sama dari semua *annotator*, maka label tersebut menjadi label emas.
- b. Apabila terdapat *annotator* yang memberikan label yang tidak sama dengan *annotator* lainnya, maka dilakukan pemilihan label emas dengan cara melihat mayoritas label yang diberikan.

- c. Apabila tidak ada mayoritas label yang diberikan oleh *annotator*, maka diberikan label ***broken***.

V. Perbaikan Teks

Teks hipotesis yang dirasa oleh *annotator* tidak tepat, salah tata bahasa, ejaan, maupun tanda baca, yang ditandai dengan rata-rata nilai kualitas terjemahan kurang dari 3, dapat dilakukan perbaikan. Pembuatan kalimat baru dapat menggunakan strategi:

- a. Penghilangan kata
Terdapat satu atau lebih kata yang dihapus dari kalimat premis.
- b. Penambahan kata
Menambahkan satu atau lebih kata dari kalimat premis. Contohnya, dapat menambahkan kata sifat, kata negasi, dan lain-lain.
- c. Perubahan leksikal
Menggantikan satu atau lebih kata dari kalimat premis dengan sinonim, antonim, hipernim, ataupun hiponim.
- d. Parafrase
- e. Perubahan struktur kalimat
Mengubah struktur kalimat, contohnya dari kalimat aktif menjadi kalimat pasif.
- f. Bahasa Jawa
Kalimat baru yang dibuat harus dalam bahasa Jawa Ngoko untuk area Jawa Timur.