

PANDUAN EVALUASI WORD ALIGNMENT

Pengantar

1. Evaluasi ini dilakukan untuk mengetahui seberapa baik performa alignment yang dilakukan oleh tool Giza++ pada korpus identik. Korpus identik ini merupakan kumpulan pasangan kalimat dalam dua bahasa yaitu Indonesia-Inggris.
2. Anotator (evaluator) pada task ini bertugas untuk memasangkan kata-kata dari bahasa Inggris-Indonesia pada pasangan kalimat yang diberikan.
3. Input yang diberikan adalah file yang berisi 200 pasang kalimat Indonesia-Inggris dengan format:

```
Sentence pair#<nomor>
<Kalimat bahasa Indonesia>
<Kalimat bahasa inggris>
Sentence pair#<nomor+1>
<Kalimat bahasa Indonesia>
<Kalimat bahasa inggris>
...
Sentence pair#200
<Kalimat bahasa Indonesia>
<Kalimat bahasa inggris>
```

Dengan setiap kalimat memiliki format:

Kalimat bahasa Indonesia

```
kata1(indeks1) kata2(indeks2) ... kataN(indeksN)
```

Kalimat bahasa Inggris

```
Kata1 ({ }) kata2 ({ }) kata3 ({ }) ... kataM ({ })
```

4. Output merupakan file yang sama, dengan kata-kata pada bahasa Inggris sudah dipasangkan dengan nomor indeks kata bahasa Indonesia (Contoh pemasangan dapat dilihat pada bagian **proses pemasangan kata**)
5. Pasangan kalimat secara sederhana dapat dicontohkan dengan kalimat input berikut:

```
Sentence pair#1
Kamu(1) seharusnya(2) belajar(3) dari(4) hal(5) tersebut(6)
NULL ({ }) You ({ }) should ({ }) have ({ }) learned ({ }) from ({ }) that ({ })
```

Tugas anotator adalah memasangkan kata-kata dalam bahasa Inggris dengan kata yang bersesuaian pada bahasa Indonesianya. Pada contoh kalimat tersebut, pasangan kata yang tepat adalah:

- “You” dengan “Kamu”

- “should” dengan “seharusnya”
- “learned” dengan “belajar”
- “from” dengan “dari”
- “that” dengan “hal tersebut”

Proses Pemasangan Kata

1. Diberikan pasangan kalimat dalam bahasa Inggris dan bahasa Indonesia sebanyak 200 buah pasangan:

Contoh **input** Pasangan 1:

Sentence pair#1
 Lalu(1) ia(2) mengalami(3) kedamaian(4) yang(5) tidak(6) terlukiskan(7)
 NULL ({}) He ({}) then ({}) experienced ({}) an ({}) indescribable ({}) peace ({})

Cara membaca:

Indeks	1	2	3	4	5	6	7
Kata	Lalu	ia	mengalami	kedamaian	yang	tidak	terlukiskan

2. Dari pasangan kalimat tersebut lakukan pemasangan **satu kata dengan kata lain** dengan cara:

2.1. Berikan angka indeks yang sesuai pada kata-kata dalam bahasa Inggris yang bersesuaian (Kata yang bersesuaian tidak harus *direct translation* tetapi bisa yang maknanya serupa).

Contoh pada pasangan 1:

- Kata “**ia**” berpasangan dengan “**he**”, maka **He ({})** diisi dengan angka 2(ia)
- Sehingga menjadi **He ({} 2)**
- Kata “**indescribable**” berpasangan dengan “**tidak**” dan “**terlukiskan**”, maka **indescribable ({})** diisi dengan angka **6(tidak)** dan **7(terlukiskan)**
- Sehingga menjadi **indescribable ({} 6 7)**

2.2. Berikan angka indeks pada token **NULL**, pada kata-kata bahasa Indonesia yang tidak memiliki pasangan.

Contoh pada pasangan 1:

- Kata “**yang**” pada kalimat tersebut tidak memiliki pasangan, sehingga diletakan pada token **NULL ({})**
- Sehingga menjadi **NULL ({} 5)**

2.3. Hasil akhir merupakan kalimat bahasa Inggris yang sudah dipasangkan dengan angka indeks tersebut.

Contoh hasil output:

Sentence pair#1
Lalu(1) ia(2) mengalami(3) kedamaian(4) yang(5) tidak(6) terlukiskan(7)
NULL ({ 5 }) He ({ 2 }) then ({ 1 }) experienced ({ 3 }) an ({ }) indescribable ({ 6 7 })
peace ({ 4 })

3. **Bila terdapat kebingungan dalam memasangkan kata**, *annotator* dapat membuka kamus Indonesia-English untuk mencari *vocabulary* yang dibutuhkan.
4. **Bila terdapat kalimat yang terpotong atau tidak selaras dengan kalimat pasangannya**, laporkan nomor kalimat tersebut untuk dapat diganti dengan kalimat lain.

Contoh kasus yang mungkin terjadi:

1. Terdapat satu kata bahasa Inggris yang mempunyai pasangan dengan lebih dari satu kata dalam bahasa Indonesia seperti:
 - a. **Hospital** dengan **rumah sakit**
 - b. **Smaller** dengan **lebih kecil**

Contoh kasus dalam pasangan kalimat:

Tidak(1) seorang(2) pun(3) dari(4) klub(5) tersebut(6) yang(7) segera(8) memberi(9)
komentar(10)
NULL ({ 4 7 9 6 }) Nobody ({ 1 2 3 }) at ({ }) the ({ }) club ({ 5 }) was ({ }) immediately ({ 8 })
available ({ }) for ({ }) comment ({ 10 })

Kata “**Tidak**”, “**seorang**”, dan “**pun**” dipasangkan dengan kata “**Nobody**”.

2. Terdapat satu kata bahasa Indonesia yang mempunyai pasangan dengan lebih dari satu kata dalam bahasa Inggris seperti:
 - a. **Menyukainya** dengan **like him**
 - b. **Menyerah** dengan **give up**

Contoh kasus dalam pasangan kalimat:

Akhirnya(1) ani(2) menyerah(3) dalam(4) lomba(5) itu(6)
NULL ({ }) ani ({ 2 }) finally ({ 1 }) gives ({ 3 }) up ({ 3 }) on ({ 4 }) that ({ 6 }) competition ({ 5 })

3. Terdapat kata-kata yang tidak memiliki pasangan pada kalimat bersangkutan seperti:

Contoh pada pasangan kalimat:

tahun(1) yang(2) lalu(3) hanya(4) sekitar(5) 4(6) sekian(7) persen(8)

NULL ({ 2 4 5 7 }) Last ({ 3 }) year ({ 1 }) it ({ }) expanded ({ }) 4 ({ 6 }) percent ({ 8 })

Kata “**yang**”, “**hanya**”, “**sekitar**”, dan “**sekian**” dipasang dengan **NULL** karena tidak ada kata yang bersesuaian pada kalimat bahasa Inggris diatas.

4. Terdapat kata-kata yang merupakan terjemahan secara implisit, seperti

Pekerjaan(1) yang(2) sangat(3) menyenangkan(4)
NULL ({ 2 }) awesome ({ 3 4 }) job ({ 1 })

5. Kasus dimana terdapat panggilan/singkatan, seperti:

Susilo(1) Bambang(2) Yudhoyono(3) turun(4) dari(5) jabatan(6)
NULL ({ 4 5 6 }) SBY ({ 1 2 3 }) is ({ }) retired ({ 4 5 6 })

Atau contoh lain

Susilo(1) Bambang(2) Yudhoyono(3) (SBY)(4) turun(5) dari(6) jabatan(7)
NULL ({ 4 5 6 }) SBY ({ 1 2 3 4 }) is ({ }) retired ({ 5 6 7 })

6. Kasus dimana terdapat tanda baca pada kalimat, seperti:

Dia(1) gagal(2) menuntaskannya!(3)
NULL ({ }) He ({ 1 }) is ({ }) failed!({ 2 }) to ({ }) finish ({ 3 }) it ({ 3 })

Pada kasus masih terdapat tanda baca di kata-kata dalam kalimat, **abaikan** tanda baca tersebut.