

LAPORAN TUGAS PEKAN 9 NLP : WORD SENSE DISAMBIGUATION

CLARISA HASYA YUTIKA | 1301174256 | IF 41 GAB01

1. Deskripsi Masalah

Implementasi algoritma Lesk untuk menentukan sense yang tepat pada sebuah kata Bahasa Indonesia. Data definisi sense dan contoh penggunaan dalam sebuah konteks diambil dari Kamus Besar Bahasa Indonesia daring, <https://kbbi.kemdikbud.go.id>. Kemudian pilih 3 kata target yang memiliki sense lebih dari 1 untuk diuji dengan algoritma Lesk, yang menghitung skor kemiripan/similarity antara 2 buah teks berdasarkan jumlah kata yang beririsan pada 2 buah teks tersebut. Kemudian membuat 2 kalimat uji dengan sense yang berbeda dari setiap kata target, dan diuji menggunakan algoritma Lesk untuk menghasilkan sense yang tepat dari kalimat uji tersebut.

2. Perancangan Sistem

• Mendefinisikan sense

word	sense	definition	examples
bisa	bi.sa (1) (v)	mampu (kuasa melakukan sesuatu); dapat	ia -- membaca, tetapi tidak -- menulis
bisa	bi.sa (2) (n) 1	zat racun yang dapat menyebabkan luka, busuk, ...	None
bisa	bi.sa (2) (n) (ki) 2	sesuatu yang buruk, yang dapat merusakkan akhl...	ajaranmu itu akan menjadi -- bagi kami
bisa	bi.sa (3) (n) (Wl)	sapaan kepada dukun yang berasal dari golongan...	None
bunga	bu.nga (1) (n) 1	bagian tumbuhan yang akan menjadi buah, biasan...	-- mangga; -- rambutan
bunga	bu.nga (1) (n) 2	jenis untuk berbagai-bagai bunga	-- melati; -- mawar
bunga	bu.nga (1) (n) 3	gambar hiasan (pada kain, pamor ukiran, dan se...	None
bunga	bu.nga (1) (n) 4	tambahan untuk memperindah	ceritanya itu sudah banyak --nya
bunga	bu.nga (1) (n) 5	tanda-tanda baik	sudah tampak --nya bahwa tuntutan kita akan be...
bunga	bu.nga (1) (n) (ki) 6	sesuatu yang dianggap elok (cantik) seperti bunga	gadis itu adalah -- di kampungnya
bunga	bu.nga (1) (n) (ki) 7	nama yang dipakai untuk beberapa jenis tumbuha...	-- bangkai, -- ketongkeng, -- manila
bunga	bu.nga (2) (n) (Ek) 1	imbalan jasa untuk penggunaan uang atau modal ...	None
bunga	bu.nga (2) (n) (Ek) 2	pendapatan atas setiap investasi modal	None
tahu	ta.hu (1) (v) 1	mengerti sesudah melihat (menyaksikan, mengala...	ia -- bahwa saya yang menolongnya; perkara mes...
tahu	ta.hu (1) (v) 2	(v) kenal (akan); mengenal	ia tidak -- akan sanak saudaranya lagi
tahu	ta.hu (1) (v) 3	mengindahkan; memedulikan	ia sudah tidak mau -- lagi kepada anaknya
tahu	ta.hu (1) (v) 4	mengerti; berpengertian	siapa yang -- apa maksud tanda ini?
tahu	ta.hu (1) (v) 5	pandai; cakap	sedikit-sedikit saya -- juga tentang mesin
tahu	ta.hu (1) (v) 6	insaf; sadar	dia tidak -- akan kekurangannya
tahu	ta.hu (1) (v) (cak) 7	pemah	petinju itu tidak -- menang, adikku tidak -- m...
tahu	ta.hu (2) (n)	makanan dari kedelai putih yang digiling halus...	None

• Function Simplified Lesk

```
def lesk( word, sentence ):
    bestsense = None
    maxoverlap = 0
    senses = thesaurus.loc[word, : ]
    for id, sense in senses.iterrows():
        if sense['examples'] != None:
            sense['definition'] = re.sub(r'\W', ' ', sense['definition'])
            sense['examples'] = re.sub(r'\W', ' ', sense['examples'])
            signature = nltk.word_tokenize(sense['definition']) + nltk.word_tokenize(sense['examples'])
        else:
            sense['definition'] = re.sub(r'\W', ' ', sense['definition'])
            signature = nltk.word_tokenize(sense['definition'])
    overlap = overlapcontext(set(signature), set(sentence))
    if overlap > maxoverlap:
        maxoverlap = overlap
        bestsense = sense['sense']
    return bestsense
```

• Function Compute Overlap

```
def overlapcontext( signature, sentence ):
    count = 0
    signature = signature.difference(stopword)
    sentence = sentence.difference(stopword)
    for word in signature:
        for word_sense in sentence:
            if (word == word_sense):
                count += 1
                # print(word, word_sense)
    return count
```

3. Hasil dan Analisis

Kalimat	Actual Sense	Prediction Sense
Saya <u>bisa</u> menulis	bi.sa (1) (v)	bi.sa (1) (v)
Kucing itu terkena <u>bisa</u> ular sehingga mati	bi.sa (2) (n) 1	bi.sa (2) (n) 1
Kamu memberi saya <u>bunga</u> mawar	bu.nga (1) (n) 2	bu.nga (1) (n) 2
Meminjam uang di bank akan terkena <u>bunga</u>	bu.nga (2) (n) (Ek) 1	bu.nga (1) (n) 2
Adik suka makan <u>tahu</u> kedelai	ta.hu (2) (n)	ta.hu (2) (n)
Ayah ternyata <u>tahu</u> atau kenal dengan dia	ta.hu (1) (v) 2	ta.hu (1) (v) 2

Dipilih 3 kata target yang memiliki lebih dari 1 sense adalah ‘bisa’, ‘bunga’, ‘tahu’. Memilih kata tersebut karena kata tersebut sering digunakan dalam sehari-hari dan banyak sekali contoh kalimat yang ambigu.

Dari hasil uji terdapat 1 kalimat yang hasil prediksi dengan algoritma Lesk tidak sesuai dengan actual sensenya. Hal ini dapat terjadi karena kurang tepatnya atau tidak ada kata-kata yang beririsan antara kalimat uji dan kalimat gloss + examples. Pada kalimat ‘meminjam uang di bank akan terkena bunga’, sense seharusnya adalah ‘bu.nga (2) (n) (Ek) 1’ dengan gloss ‘imbalan jasa untuk penggunaan uang atau modal yang dibayar pada waktu tertentu berdasarkan ketentuan atau kesepakatan, umumnya dinyatakan sebagai persentase dari modal pokok’. Tetapi hasil prediksi dengan algoritma Lesk adalah sense ‘bu.nga (1) (n) 2’. Karena pada gloss + examples pada sense tersebut, terdapat kata ‘bunga’ sedangkan tidak ada irisan terhadap gloss + examples dari sense ‘bu.nga (2) (n) (Ek) 1’ sehingga kalimat tersebut terprediksi dengan sense ‘bu.nga (1) (n) 2’.

Kemudian pada kalimat lainnya, hasil prediksi sense nya sesuai dengan actual sense nya karena terdapat irisan kata yang tepat antara 2 kalimat tersebut.