

1. Berikan analisis dugaan Anda mengapa tipe question ABBR kinerjanya paling rendah (sulit dideteksi), sementara tipe question LOC mempunyai kinerja paling tinggi?

Tipe question ABBR sulit dideteksi karena biasanya di awal kalimat pertanyaan tipe ABBR adalah “what”, sedangkan tipe question “ENTY”, “DESC” juga biasanya diawali dengan “what”. Kata “abbreviation”, “stand for” terkadang terdapat dalam pertanyaan tipe “ABBR”, tetapi tidak selalu. Contoh “*What does the word LASER mean ?*” pertanyaan tersebut dapat dimaksud “ABBR” atau “DESC”.

Tipe question “LOC” lebih mudah dideteksi karena pertanyaan yang mengandung kata “where” pasti menunjukkan lokasi.

2. Apakah menurut Anda dapat diterapkan metode gabungan dengan pendefinisian aturan secara manual untuk mendeteksi satu atau beberapa tipe question? Sebutkan contoh aturan yang dapat diterapkan untuk mendeteksi tipe question (dalam bentuk pseudocode/algoritma).

```
if (("who" or "whom") in current_question):  
    current_label = "HUM"  
elif ("where" in current_question):  
    current_label = "LOC"  
elif (("How much" or "How many") in current_question):  
    current_label = "NUM"  
elif (("abbreviation" or "stand for") in current_question):  
    current_label = "ABBR"
```

Dapat diterapkan metode gabungan dengan pendefinisian manual pada beberapa tipe question. Aturan tersebut diterapkan sebelum proses klasifikasi supervised.

Pada tipe question “ENTY” dan “DESC” sulit untuk dideteksi, karena diawal kalimat pertanyaan tipe tersebut biasanya diawali dengan kata “what”, sehingga tidak terdapat suatu ciri khas yang dapat membedakan tipe question tersebut.

Pada tipe question “HUM” terdapat 2 kata yaitu “who” dan “whom” yang pasti menggambarkan bahwa kata tersebut bertuju kepada seseorang.

Pada tipe question “LOC”, kata “where” pasti bertujuan untuk lokasi, tempat, wilayah.

Pada tipe question “NUM”, dengan pertanyaan “how much” dan “how many” pasti akan menghasilkan jawaban numerik.

Pada tipe question “ABBR”, pada awal kalimat pertanyaan biasanya diawali dengan kata “what”, tetapi kata tersebut memiliki banyak makna. Sehingga tidak digunakan dalam rule. Kemudian kata yang menjadi ciri khas nya adalah “abbreviation” dan “stand for” dalam kalimat pertanyaan. Tetapi walaupun terdapat ciri khas tersebut, terdapat beberapa pertanyaan yang tidak mengandung kedua kata tersebut. Sehingga untuk tipe question “ABBR” masih sulit untuk dideteksi.

LAPORAN TUGAS PEKAN 14 NLP : QUESTION ANSWERING SYSTEM

CLARISA HASYA YUTIKA | 1301174256 | IF 41 GAB01

3. Terapkan aturan tambahan yang didefinisikan secara manual tersebut untuk mengidentifikasi tipe question pada dataset uji yang sama, lalu amati hasil evaluasinya!

- Sebelum ditambah aturan

	precision	recall	f1-score	support
ABBR	0.00	0.00	0.00	4
DESC	0.71	0.54	0.62	37
ENTY	0.52	0.67	0.59	51
HUM	0.78	0.76	0.77	50
LOC	0.85	0.88	0.86	32
NUM	0.80	0.77	0.78	26
accuracy			0.70	200
macro avg	0.61	0.60	0.60	200
weighted avg	0.70	0.70	0.70	200

- Sesudah ditambah aturan

	precision	recall	f1-score	support
ABBR	0.00	0.00	0.00	4
DESC	0.72	0.57	0.64	37
ENTY	0.52	0.67	0.59	51
HUM	0.78	0.76	0.77	50
LOC	0.85	0.88	0.86	32
NUM	0.83	0.77	0.80	26
accuracy			0.70	200
macro avg	0.62	0.61	0.61	200
weighted avg	0.71	0.70	0.70	200

Tidak terdapat perbedaan yang signifikan antara sebelum dan sesudah ditambah aturan. Dari label yang sudah ada dengan hasil label dengan aturan yang didefinisikan manual hanya sedikit perbedaan, dan dari aturan yang sudah didefinisikan, tidak terdapat aturan untuk tipe question “DESC” dan “ENTY”, sedangkan pada dataset terdapat banyak tipe question tersebut. Sehingga aturan yang sudah didefinisikan tidak terlalu berpengaruh untuk identifikasi tipe question.

4. Kira-kira fitur apa lagi yang dapat ditambahkan untuk meningkatkan kinerja identifikasi tipe question? Lengkapi jawaban poin ini dengan referensi.

Question types	Features
How, Who, Why, When, Where, Which	Unigram, Bigram, Word-Shapes, Question Pattern
What	Unigram, Bigram, Head word, Word-Shape, Related Words, Question Pattern, Query Expansion, Question Category
Other questions	Unigram, Bigram, Word-Shape, Related Words

Fitur ekstraksi yang dapat dilakukan adalah dengan TF-IDF, N-gram, word-shape, dsb. Berdasarkan paper “Improving Question Classification by Feature Extraction and Selection” oleh Van-Tu Nguyen dan Len Anh-Cuong. Pada paper tersebut, tahapan yang dilakukan adalah feature selection, kemudian desain jenis fitur baru berdasarkan pola pertanyaan. Setelah itu untuk proses klasifikasi menggunakan Support Vector Machines (SVM). Hasil yang diperoleh dengan akurasi 95,2% dan 91,6%.

Reference :

Van-Tu, N., & Anh-Cuong, L. (2016). Improving question classification by feature extraction and selection. Indian Journal of Science and Technology, 9(17), 1-8.