LAPORAN TUGAS BESAR

LEXICAL ANALYZER DAN PARSER

Disusun untuk memenuhi tugas besar mata kuliah Teori Bahasa dan Automata

Dosen Pengampu: Ahmad Suryan, M.T.

Oleh Kelompok 8:

Naufal Abdurrahman Burhani (1301204008)
 Harvian Khusnan Hafidz (1301204303)
 Khalillullah Al Faath (1301204376)



Fakultas Informatika
Telkom University
2022

Daftar Isi

Daftar Isi		2
1. Tinjau	an Pustaka	4
1.1. Ba	hasa Rusia dan Aksara Kiril	4
1.2. <i>Le</i>	xical Analysis	5
1.2.1.	Token	5
1.2.2.	Deterministic Finite Automata (DFA)	5
1.2.3.	Context-free Grammar	5
1.3. Pa	rser	6
1.3.1.	LL(k) Parser	6
1.3.2.	LR(k) Parser	6
1.3.3.	LL(k) Parser vs LR(k) Parser	6
2. Ketent	uan Program	6
2.1. Co	ontext-free Grammar	6
2.2. Fi	nite Automata	7
2.2.1.	FA Subject	7
2.2.2.	FA Verb	8
2.2.3.	FA Object	8
2.2.4.	FA Gabungan	9
2.3. Pa	rse table	10
3. Impler	nentasi Program	10
3.1. Va	nriabel-variabel dalam pembangunan web	10
3.2. Er	umerasi list alfabet dan daftar state	11
3.3. Da	aftar Transition Table	12
3.4. Le	xical Analyzer	13
3.5. Da	aftar symbol terminal dan nonterminal serta inisialisasi parse table	13
3.6. Da	aftar Parser Table	14
4. Hasil l	Program	16
4.1. Ta	mpilan web	16
4.1.1.	Lexical Analyzer	16
4.1.2.	Parser	16

4.2. Ou	utput program	16
4.2.1.	Contoh input valid lexical analyzer	16
4.2.2.	Contoh input invalid lexical Analyzer	17
4.2.3.	Contoh input valid parser	19
4.2.4.	Contoh input invalid parser	21
5. Resour	rces dan tutorial	23
5.1. Re	esources	23
5.2. Tu	utorial	23
6. Bibliog	graphy	25
_		

1. Tinjauan Pustaka

1.1. Bahasa Rusia dan Aksara Kiril

Bahasa Rusia adalah bahasa yang termasuk dalam rumpun bahasa Slavia Timur, yang merupakan bahasa yang umumnya digunakan di negara-negara bekas Uni Soviet. Bahasa Rusia juga termasuk ke dalam bahasa resmi PBB (Esipova & Gradirovski, 2008).

Bahasa Rusia umumnya memakai aksara Kiril (bahasa Inggris: cyrillics, rusia: кириллица). Saat ini, alfabet Kiril yang dipakai di Rusia telah dimodifikasi sedemikian rupa dan terdiri atas sebelas huruf vokal, dua puluh konsonan, serta dua penanda keras dan lunak sehingga seluruhnya ada 33 huruf.

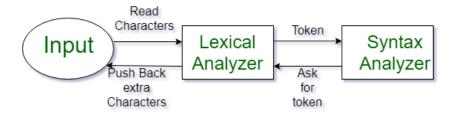
Berikut adalah aksara dalam bahasa Rusia.

Huruf	Nama huruf	Alih Aksara	Alih Aksara		
	Nama nurui	Allii Aksara	nama huruf		
Аа	а	а	а		
Бб	бэ	b	be		
Вв	вэ	V	ve		
Гг	гэ	g	ge		
Дд	дэ	d	de		
Еe	йэ	ye	ye		
Ëë	йо	yo	уо		
Жж	жэ	zh	zhe		
3 3	39	z	ze		
Ии	И	i	i		
Йй	и краткое	у	i kratkoye		
Кк	ка	k	ka		
Лл	эл	T	el		
Мм	эм	m	em		
Нн	эн	n	en		
0 0	0	0	0		
Пп	пэ	р	pe		
Рр	эр	r	er		
Сс	эс	S	es		
Тт	тэ	t	te		
Уу	у	u	и		
Фф	эф	f	ef		
Хx	ха	h	ha		
Цц	цэ	ts	tse		
Чч	че	ch	che		
Шш	ша	sh	sha		
Щщ	ща	sch	scha		

Ъъ	твёрдый знак	*	tvyordiy znak
Ыы	Ы	у	У
Ьь	мя́гкий знак	*	myagkiy znak
Ээ	э	е	е
Юю	йу	yu	yu
Яя	йа	ya	ya

1.2. Lexical Analysis

Lexical Analysis adalah tahap pertama dalam compiler yang dengannya program input tingkat tinggi diubah menjadi deretan token. Lexical Analysis ini dapat diimplementasikan dengan menggunakan Deterministic Finite Automata (DFA).



Outputnya adalah deretan token yang kemudian akan dikirimkan ke parser untuk dilakukan analisis sintaks.

1.2.1. Token

Token adalah kumpulan karakter yang dapat diperlakukan sebagai sebuah unit dalam grammar suatu bahasa (GeeksforGeeks, 2021).

1.2.2. Deterministic Finite Automata (DFA)

DFA (Deterministic Finite Automata) adalah Finite-state Machine atau mesin keadaan terbatas yang menerima atau menolak string dari simbol dan hanya menghasilkan perhitungan unik dari otomata untuk setiap string yang di masukan (Belajar CPP, 2018).

1.2.3. Context-free Grammar

Context-free Grammar (CFG) adalah sebuah grammar yang aturan produksinya berbentuk $A \to \alpha$, di mana A adalah sebuah symbol non-terminal dan α adalah kumpulan symbol terminal dan non-terminal dengan catatan α bisa merupakan string kosong (John E. Hopcroft, et.al., 2001).

1.3. Parser

Jenis Parser ada 2:

- LL(k) parser
- LR(k) parser

1.3.1. LL(k) Parser

Alasan dinamakan LL parser adalah:

- L pertama menunjukkan string input akan dibaca dari kiri menuju ke kanan.
- L kedua menandakan prase tree yang dibangun akan menghasilkan leftmost derivation

1.3.2. LR(k) Parser

Alasan dinamakan LR parser adalah:

- L menunjukkan string input akan dibaca dari kiri menuju ke kanan.
- R menandakan prase tree yang dibangun akan menghasilkan rightmost derivation.

1.3.3. LL(k) Parser vs LR(k) Parser

- LL(k) parser dikenal dengan nama top-down parser. Dimulai dengan Start simbol pada stack dan secara berulang memecah non terminal pada stack ke dalam komponennnya sampai string simbol cocok dengan input string yang digenerate.
- LR(k) parser diklasifikasikan sebagai bottom-up parser. Dimulai dari aktivitas bersamaan untuk membentuk kejadian non terminal dari symbolsymbol pada string sampai ditemukan start simbol dari grammar yang men-generate.

2. Ketentuan Program

2.1. Context-free Grammar

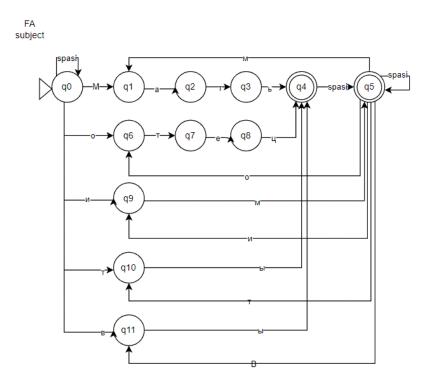
Notasi yang digunakan untuk context-free grammar adalah Backus-Naur form. Berikut ini adalah context-free grammar yang digunakan dengan notasi Backus-Naur form.

```
<$> ::= <$B> <VB> <OB>
<$B> ::= "мать" | "отец" | "им" | "ты" | "вы"
<\VB> ::= "готовит" | "уяить" | "тянуть" | "рисовать" | "читать"
<\OB> ::= "книга" | "комар" | "ром" | "ученик"
```

2.2. Finite Automata

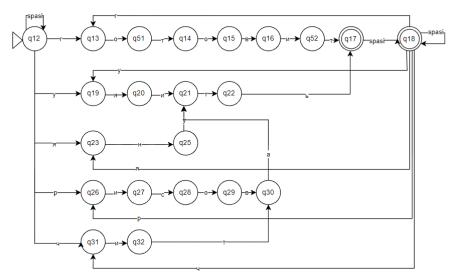
Setelah membuat aturan untuk grammar, langkah selanjutnya yang akan dilakukan adalah membangun finite automata sebagai rules state-state serta value yang akan digunakan pada program lexical analyzer yang akan dibuat. Adapun bentuk model finite automata yang dibuat adalah berikut ini.

2.2.1. FA Subject



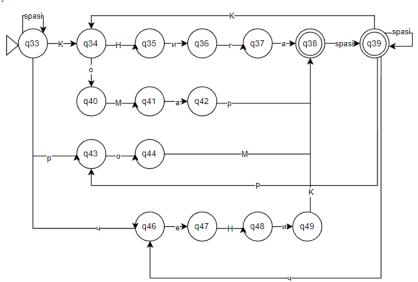
2.2.2. FA Verb



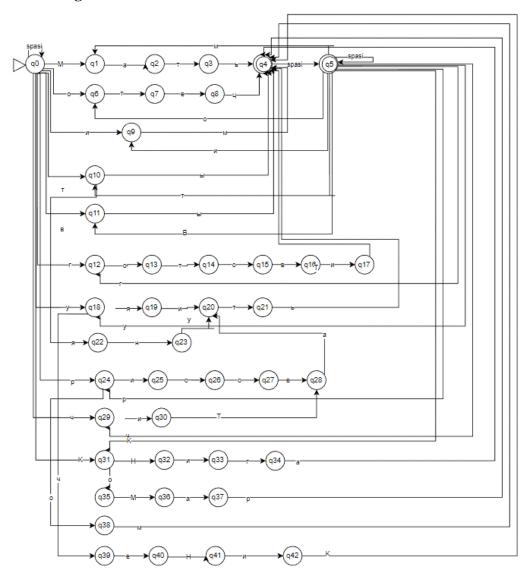


2.2.3. FA Object





2.2.4. FA Gabungan



2.3. Parse table

	мать	отец	ИМ	ТЫ	ВЫ	готовит	уяить	рисовать	читать	книга	комар	ром	ученик	EOS
	SB	SB	SB	SB	SB									
S	VB	VB	VB	VB	VB	error	error	error	error	error	error	error	error	error
	OB	OB	OB	OB	OB									
SB	мать	отец	ИМ	ты	ВЫ	error	error	error	error	error	error	error	error	error
VB	error	error	error	error	error	готовит	уяить	рисовать	читать	error	error	error	error	error
ОВ	error	error	error	error	error	error	error	error	error	книга	комар	ром	ученик	error

3. Implementasi Program

Program yang dibuat adalah program berbasis web dengan menggunakan bahasa Javascript. Untuk pembangunan websitenya menggunakan HTML dan CSS.

3.1. Variabel-variabel dalam pembangunan web

```
var inputKalimat = document.getElementById('input_kalimat')
;
var hasil = document.getElementById('result');
var clear = document.getElementById('btn-clear');
var loading = document.getElementById('loading');
var form = document.getElementById('form')
```

3.2. Enumerasi list alfabet dan daftar state

Karena aksara yang digunakan bukan aksara latin maka perlu mengambil Unicodenya. Untuk aksara Kiril dalam Unicode terdiri dari decimal 1024 sampai 1279. Untuk menghandle huruf latin maka ditambahkan ke dalam list dengan Unicode dari 65—122.

Dilakukan juga pembentukan state yang terdiri dari "q" yang diikuti angka dari 0—24.

```
var alfabet = [];
for (let i = 1024; i <= 1279; i++){
    let text = String.fromCharCode(i);
    alfabet.push(text);
}

for (let i = 65; i <= 122; i++){
    let text = String.fromCharCode(i);
    alfabet.push(text);
    You, 27 seconds ago * fix *
}

// const alfabet = /^[\u0400-\u04FF]+$/;
// console.log('\u0400-\u04FF]+$/;
// console.log('Hello:', alfabet.test('\u0400-\u04FF]);

var stateList = [];

let N = 42;
for (let i = 0; i <= N; i++){
    var strAngka = i.toString();
    var str = "q";
    var namaState = str.concat(strAngka);
    stateList.push(namaState)
}</pre>
```

3.3. Daftar Transition Table

Menginisialisasi untuk setiap alfabet sebagai error. Setelah itu dibuat transition table sesuai dengan FA gabungan.

```
for(var state in stateList) {
    for(daftarAlfabet in alfabet) {
        transitionTable[[stateList[state], alfabet
        [daftarAlfabet]]] = 'ERROR'
    }
    transitionTable[[stateList[state], '#']] = 'ERROR'
    transitionTable[[stateList[state], ' ']] = 'ERROR'
}
```

```
// initial state
transitionTable[["q0"," "]] = "q0";

// final state
transitionTable[["q4"," "]] = "q5";
transitionTable[["q4"," "]] = "ACCEPT";
transitionTable[["q5"," "]] = "q5";
transitionTable[["q5"," "]] = "ACCEPT";

//Subject Mamb
transitionTable[["q0"," M"]] = "q1"
transitionTable[["q1"," a"]] = "q2"
transitionTable[["q2"," T"]] = "q3"
transitionTable[["q3"," b"]] = "q4"

transitionTable[["q5"," M"]] = "q6"
transitionTable[["q6"," T"]] = "q6"
transitionTable[["q6"," T"]] = "q8"
transitionTable[["q8"," "]] = "q4"

transitionTable[["q8"," "]] = "q4"
transitionTable[["q8"," "]] = "q6"
```

3.4. Lexical Analyzer

Dibuat lexical analyzer mengikut pada tata cara analisis leksikal pada poin 1.

```
form.onsubmit = (event) => {
    event.preventDefault()
    loading.style = 'display: inline-block'
   var indexChar = 0;
   var state = 'q0';
   var currentToken = '';
   var validation = '';
   var inputChar = inputKalimat.value + '#';
    console.log(inputChar);
    while (state != 'ACCEPT') {
       var currentChar = inputChar.charAt(indexChar)
       currentToken += currentChar
       state = transitionTable[[state, currentChar]]
        if(state == 'q4') {
           console.log("valid gais")
           validation += "valid "
           currentToken = ''
        if(state == 'ERROR') {
           console.log("ERROR")
           validation += "ERROR "
```

3.5. Daftar symbol terminal dan nonterminal serta inisialisasi parse table

```
var nonTerminals = ['S', 'SB','VB','OB'];
var terminals = ['мать', 'отец','им', 'ты','вы','готовит',
'уяить','рисовать',
'читать','книга','комар', 'ром',
'ученик'];
var parse_table = [];
```

3.6. Daftar Parser Table

```
parse_table[['S','мать']] = ['SB', 'VB', 'ОВ']
parse_table[['S','отец']] = ['SB', 'VB', 'OB']
parse_table[['S','им']] = ['SB', 'VB', <u>'</u>OB']
parse_table[['S', 'ты']] = ['SB', 'VB', 'OB']
parse_table[['S','вы']] = ['SB','VB','ОВ']
parse_table[['S','готовит']] = ['error']
parse_table[['S','уяить']] = ['error']
parse_table[['S','рисовать']] = ['error']
parse_table[['S','читать']] = ['error']
parse_table[['S','книга']] = ['error']
parse_table[['S','комар']] = ['error']
parse_table[['S','pom']] = ['error']
parse_table[['S','ученик']] = ['error']
parse_table[['S','EOS']] = ['error']
parse table[['SB','мать']] = ['мать']
parse table[['SB','отец']] = ['отец']
parse_table[['SB','um']] = ['um']
parse_table[['SB', 'ты']] = ['ты']
parse_table[['SB','вы']] = ['вы']
parse_table[['SB','готовит']] = ['error']
parse_table[['SB','уяить']] = ['error']
parse_table[['SB','рисовать']] = ['error']
parse_table[['SB','читать']] = ['error']
parse table[['SB','книга']] = ['error']
parse table[['SB','комар']] = ['error']
parse_table[['SB','pom']] = ['error']
parse_table[['SB','ученик']] = ['error']
parse_table[['SB','EOS']] = ['error']
```

```
parse_table[["VB",'мать']] = ["error"]
parse_table[["VB",'отец']] = ["error"]
parse_table[["VB",'им']] = ["error"]
parse_table[["VB", 'ты']] = ["error"]
parse_table[["VB",'вы']] = ["error"]
parse_table[["VB",'готовит']] = ['готовит']
parse_table[["VB",'уяить']] = ['уяить']
parse_table[["VB",'рисовать']] = ['рисовать']
parse_table[["VB",'читать']] = ['читать']
parse_table[["VB",'книга']] = ['erroe']
parse_table[["VB",'комар']] = ['error']
parse_table[["VB",'pom']] = ['error']
parse_table[["VB",'ученик']] = ['error']
parse_table[["VB",'EOS']] = ['error']
parse_table[["ОВ",'мать']] = ["error"]
parse_table[["ОВ",'отец']] = ["error"]
parse_table[["ОВ",'им']] = ["error"]
parse_table[["ОВ", 'ты']] = ["error"]
parse_table[["ОВ",'вы']] = ["error"]
parse_table[["ОВ",'готовит']] = ['error']
parse_table[["ОВ",'уяить']] = ['error']
parse_table[["ОВ",'рисовать']] = ['error']
parse_table[["ОВ", 'читать']] = ['error']
parse_table[["ОВ",'книга']] = ['книга']
parse_table[["OB",'комар']] = ['комар']
parse_table[["OB",'pom']] = ['pom']
parse_table[["ОВ",'ученик']] = ['ученик']
parse_table[["OB",'EOS']] = ['error']
```

4. Hasil Program

4.1. Tampilan web

4.1.1. Lexical Analyzer

	Kelompok Russia	
		Kata-Kata inputan Subjek yang dapat diterima : '
	Lexical Analyzer русский язык Bahasa Rusia	
Input kata atau kalimat dalam aksara russia		
Input kalimat		
ех: ученик		
Analyze		
Hasil		
Clear		

4.1.2. Parser

Parser русский язык Bahasa Rusia						
Input kalimat dalam bahasa Rusia; kalimat yang diterima terdiri atas subjek predikat objek						
Input kalimat						
ех:мать готовит книга						
Analyze						
Hasil						
Clear						

4.2. Output program

4.2.1. Contoh input valid lexical analyzer

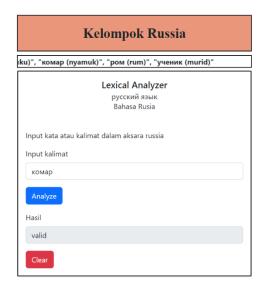
Dilakukan pengeceken dengan input sebagai berikut:

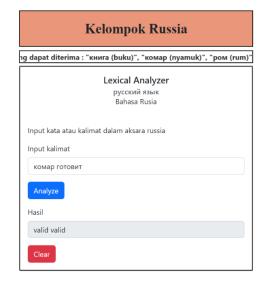
- Мать: mengeluarkan outuput "valid" karena ada pada grammar.
- Уяить: mengeluarkan output "valid" karena ada pada grammar.

- Комар: mengeluarkan output "valid" karena ada pada grammar.
- Комар готовит: mengeluarkan output "valid valid" karena kedua kata ada pada grammar







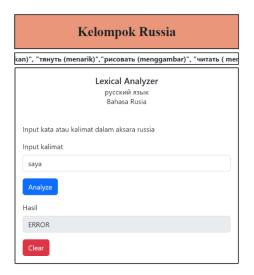


4.2.2. Contoh input invalid lexical Analyzer

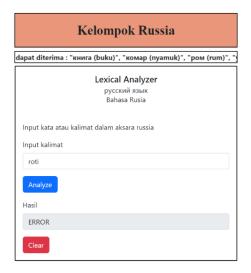
Dilakukan pengecekan input sebagai berikut.

- saya: mengeluarkan output "error" karena tidak ada pada grammar
- makan: mengeluarkan output "error" karena tidak ada pada grammar

- roti: mengeluarkan output "error" karena tidak ada pada grammar
- saya makan roti: mengeluarkan output "error" karena ketiganya tidak ada pada grammar





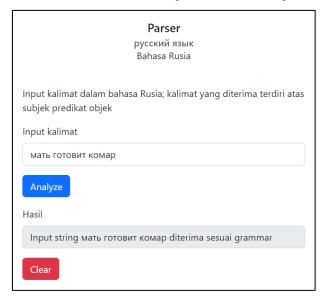




4.2.3. Contoh input valid parser

Dilakukan pengecekan input мать готовит комар:

- Мать: ada dalam grammar dan merupakan subjek.
- Готовит: ada dalam grammar dan merupakan verb.
- Комар: ada dalam grammar dan merupakan objek.
- Kalimat terdiri atas subjek, verb, dan objek, maka ia diterima.



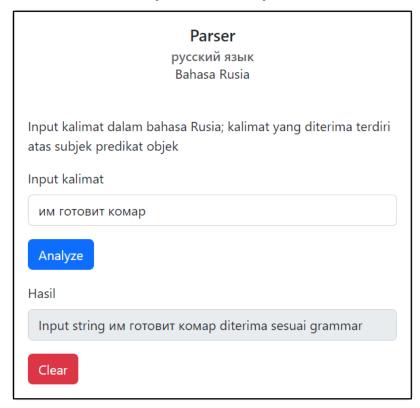
Dilakukan pengecekan input отец рисовать ром:

- отец: ada dalam grammar dan merupakan subjek.
- рисовать: ada dalam grammar dan merupakan verb.
- ром: ada dalam grammar dan merupakan objek.
- Kalimat terdiri atas subjek, verb, dan objek, maka ia diterima.



Dilakukan pengecekan input им готовит комар

- им: ada dalam grammar dan merupakan subjek.
- готовит: ada dalam grammar dan merupakan verb.
- комар: ada dalam grammar dan merupakan objek.
- Kalimat terdiri atas subjek, verb, dan objek, maka ia diterima.



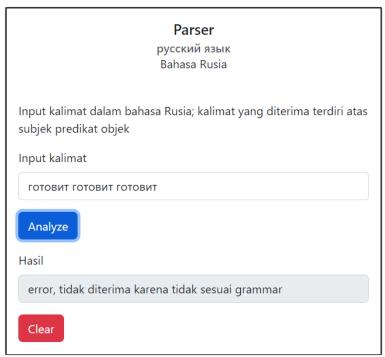
4.2.4. Contoh input invalid parser

Dilakukan pengecekan inputan мать мать мать: Keluaran error, karena kata "мать мать" tidak sesuai dengan grammar

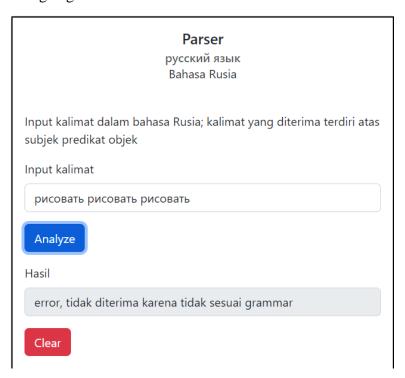


Dilakukan pengecekan inputan готовит готовит:

Keluaran error, karena kata "готовит готовит готовит" tidak sesuai dengan grammar



Dilakukan pengecekan inputan рисовать рисовать рисовать: Keluaran error, karena kata "рисовать рисовать рисовать" tidak sesuai dengan grammar



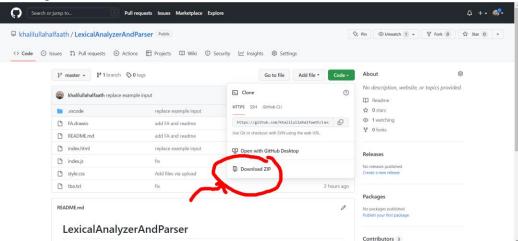
5. Resources dan tutorial

5.1. Resources

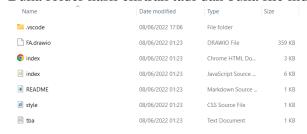
Link Github: https://github.com/khalilullahalfaath/LexicalAnalyzerAndParser

5.2. Tutorial

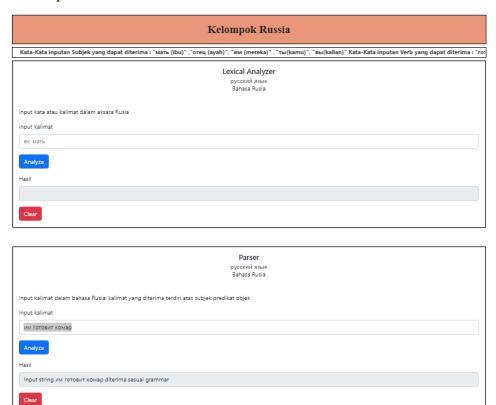
1. Download file dari link github pada resources (Jika sudah ada file zip, lompat ke Langkah 3)



- 2. Ekstrak file .zip dengan winrar atau apapun itu
- 3. Buka folder hasil ekstrak tadi dan buka file index.html



4. Maka tampilan web akan muncul



- 5. Masukkan input sesuai yang diinginkan, contoh input yang diterima ada di teks berjalan pada web
- 6. Contoh inputan valid dan tidak valid ada pada laporan bab 4 (Output program)

6. Bibliography

Esipova, S., & Gradirovski, N. (2008, August 1). *Russian Language Enjoying a Boost in Post-Soviet States*. Retrieved from Gallup.com: https://news.gallup.com/poll/109228/russian-language-enjoying-boost-postsoviet-states.aspx