

# SISTEM ENJIN PENCARIAN PANTUN DENGAN MENGUNAKAN KATA KEKUNCI

Muhammad Yasir bin Azman

Prof. Dr. Shahrul Azman Mohd Noah

*Fakulti Teknologi & Sains Maklumat, Universiti Kebangsaan Malaysia*

## ABSTRAK

Setiap pantun mempunyai maksud yang tersirat yang sukar untuk difahami jika dengan hanya melihat maksudnya secara literal. Penterjemahan bagi setiap sebuah perkataan tidak dapat untuk memastikan maksudnya yang tepat. Projek ini bertujuan untuk memudahkan pengguna dalam mengetahui makna dan mesej tersirat yang ingin disampaikan. Masalah yang akan berhadapan oleh pengguna ialah memastikan tema sebenar sesebuah rangkap pantun tersebut. Terdapat juga sebilangan pengguna merasakan pembayang yang wujud dalam rangkap tidak memberikan gambaran mesej pantun tersebut malah merasakan tiada kaitan secara langsung dengan mesej pantun. Solusi bagi permasalahan ini ialah menggunakan kaedah indeks untuk menetapkan setiap tema yang wujud dalam sesebuah pantun. Sebilangan pantun mempunyai lebih daripada satu mesej yang ingin disampaikan maka pencarian maklumat ini akan mengenalpasti tema yang wujud dalam pantun tersebut. Terdapat algoritma yang juga digunakan di dalam pantun ini adalah untuk menetapkan kedudukan pantun mengena dengan apa yang diinginkan pengguna dan terdapat juga beberapa pilihan lain yang juga disusun mengikut mana yang paling terhampir. Akhirnya dengan pencarian maklumat berasaskan kata kunci ini akan mengeluarkan pantun yang terhampir dengan kehendak pengguna dan sekaligus menyenaraikan pantun sampingan yang berkaitan.

## PENGENALAN

Minta daun diberi daun,

Dalam daun buah bidara,

Minta pantun diberi pantun,

Dalam pantun ada bicara.

(Nik Safiah Karim, 1996:15)

Pantun merupakan sebuah artifak bukan-benda (*non-material*) yang bukan sahaja digunakan dalam kehidupan orang Melayu, tetapi mempunyai berbagai-bagai makna dan fungsi dalam budayanya. Pantun sebenarnya mencerminkan peradaban Melayu dalam mana minda sebagai terjemahan perkataan inggeris '*mind*' sebahagian daripadanya. (Mohd. Taib Osman dalam Wan Abd. Kadir Wan Yusuf, 1996:1)

Pantun boleh dianggap sebagai pernyataan tradisi budaya Melayu yang menerangkan kemantapan daya kreativiti dan pemikiran orang-orang Melayu. Selain pantun, orang Melayu meluahkan buah fikiran mereka dengan pelbagai cara antaranya seni seloka, syair, gurindam

dan puisi. Orang melayu berbalas pantun sesama mereka adalah untuk menyampaikan nasihat sosialisasi dan pendidikan, menyatakan rasa rindu dan sayang, menyatakan penghargaan dan terima kasih, hiburan dan jenaka, berteka-teki menguji ketangkasan fikiran dan mengajar erti budi pekerti dan moral. Pantun mempunyai bahasa yang tersirat kerana tingginya elemen berbudi bahasa dalam kalangan masyarakat melayu. Walaupun mereka berhasrat untuk menyindir, tetapi mereka menyindir dengan ayat dan rangkap yang tersirat dan bahasa yang indah.

Bagi mereka yang tidak terbiasa dengan bahasa yang digunakan pemantun adalah sukar bagi mereka untuk memahami segala perkara yang ingin disampaikan oleh pemantun kerana laras bahasanya yang tinggi. Penggunaan perkataan di dalam pantun kadang kala bukan bermaksud sebenar perkataan tersebut malah penggunaannya merupakan perumpamaan bagi sesuatu keadaan yang lain. Perkataan pantun itu pada mulanya dipakai orang dengan makna seperti atau umpama jelas Zainal Abidin Ahmad dalam bukunya Ilmu Mengarang Melayu. Terdapat juga pendapat lain yang mengatakan perkataan pantun itu berasal daripada bahasa Minangkabau iaitu panuntun yang bermaksud pembimbing atau penasihat yang berasaskan sastera lisan dalam pengucapan pepatah dan petiti yang terkenal dalam masyarakat tersebut.

Pantun terbahagi kepada dua bahagian iaitu pembayang dan maksud pantun. Pembayang dalam setiap rangkap adalah separuh pertama daripada keseluruhan jumlah baris dalam rangkap berkenaan. Separuh kedua dalam setiap rangkap pantun pula ialah maksud atau isi pantun. Antara ciri-ciri lain yang ada dalam sebuah pantun ialah pembayangnya mempunyai hubungan yang rapat dengan alam dan persekitaran yang rapat dengan pengucapnya seperti yang terdapat dalam contoh pantun dua kerat berikut:

Sebab pulut santan terasa,

Sebab mulut badan binasa.

Pantun terdapat banyak jenis iaitu pantun dua kerat, pantun empat kerat, pantun enam kerat, pantun lapan kerat, pantun sepuluh kerat, pantun dua belas kerat, pantun empat belas kerat, pantun berkait. Pantun yang sering digunakan merupakan pantun empat kerat kerana mudah tetapi mempunyai makna yang padat. Kebiasaannya masyarakat berbalas pantun semasa majlis umum, majlis perkahwinan dan pertunangan, teks pengacara dan ucapan dalam majlis rasmi selain daripada acara pertandingan berbalas pantun. Pantun juga diterap di dalam

jiwa anak-anak bangsa ini di dalam alunan lagu rakyat seperti lagu kanak-kanak, lagu-lagu asli dan juga dikir barat.

## **PENYATAAN MASALAH**

Masyarakat kini kurang membudayakan penggunaan pantun dalam kehidupan seharian mereka. Maka, sistem ini menjadi alternative bagi melangsungkan pencarian pantun. Anantara perkara yang menjadi permasalahan sistem ini adalah sejauh mana ketepatan dan kejituan oleh kerana beberapa aspek yang mengekangnya.

Pertama, pencarian pantun sukar kerana tiada sistem yang terkini dapat menyokong pencarian pantun ini secara sistematik dan penggunaan tema tiada dimanfaatkan. Dua aspek ini menjadi antara aspek yang penting dalam memastikan pantun yang dikeluarkan mirip dengan kehendak pengguna.

Kedua, pantun melibatkan teks yang sangat terhad dan sedikit. Tambahan pula terdapat dua rangkap awalnya yang dipanggil pembayang tidak langsung menggambarkan literal maksud pantun yang terdapat pada dua rangkap yang seterusnya. Pantun ini tidak seperti dokumen dan artikel biasa. Maka, perwakilannya menjadi sangat terhad dan hanya tertumpu kepada terma atau perkataan yang wujud di dalam pantun itu. Di sini, peranan kata kekunci penting bagi mengelaspasti sinonim bagi perkataan yang bersamanya wujud di dalam pantun.

Ketiga, mengenalpasti hasil yang relevan dikeluarkan oleh sistem sangat terhad kerana kompleksiti sistem memerlukan sokongan ontology dan pemahaman rangkap bagi memastikan pantun yang dipilih sentiasa relevan.

## **OBJEKTIF KAJIAN**

Objektif pembangunan projek ini adalah

- i. Membangunkan antara muka pencarian maklumat bagi mencari pantun yang sesuai dengan tema pantun yang pengguna mahukan berdasarkan kueri yang dimasukkan pengguna ke dalam sistem.
- ii. Membangun sistem pencarian pantun dengan penggunaan *indexing*.

- iii. Mengumpulkan sebanyaknya pantun yang ada untuk dijadikan hasil kepada pencarian sistem ini.

## **METOD KAJIAN**

Dalam pembinaan sistem pencarian pantun ini, sistem ini menggunakan python dalam menyendalikan setiap perkara dalam enjin pencarian ini. Pada awalnya pantun ini dikumpulkan dan diletakan di dalam suatu pangkalan data. Setiap pantun itu kemudiannya dikategorikan mengikut jenis seperti pantun dua kerat, pantun empat kerat, pantun enam kerat dan lain-lain lagi. Kemudian pantun ini akan ditetapkan index pada kata kekunci yang berkaitan dengan makna asal pantun ini. Pantun ini pertamanya dikumpul dan maksud atau temanya dikenalpasti bagi tujuan memudahkan pembuatan indexing. Setiap pantun diletakkan di dalam format yang berasingan. Apabila sistem ini digunakan setiap format ini akan dilalui oleh sistem dan peroses penganalisaan jenis pantun akan dibuat bergantung kepada index yang telah ditetapkan.

Sistem ini akan pertamanya mencari pantun berdasarkan kata kekunci yang pengguna masukkan. Proses seterusnya, kata kekunci itu akan dia perkataannya di dalam index yang tersedia. Bagi situasi yang paling ideal, kekunci yang pengguna masukkan mengena dan serupa secara tepat. Kemudian pantun yg kata kekuncinya serasi dengan index pantun tersebut, pantun itulah yang akan dipaparkan dan dipilih untuk dihantar kepada pengguna.

Jika kata kekuncinya tidak sama secara tepat proses seterusnya akan berjalan iaitu proses pengubahan kata kekunci kepada kata dasar bagi perkataan tersebut dengan menyingkirkan segala penambahan imbuhan awalan atau akhiran. Kemudian proses yang sama dilalui dengan membandingkan kata kekunci yang sudah diubah tersebut dengan index yang ditetapkan di dalam pangkalan data pantun itu. Pantun yang mempunyai index yang sama akan dipilih dan akan dihantar kepada pengguna. Jika proses kedua ini juga gagal mencari perbandingan yang tepat dengan kata kekunci itu, terdapat satu proses pencarian perkataan yang mempunyai makna yang sama dengan makna kata kekunci tersebut atau disebut sebagai sinonim bagi perkataan itu. Seterusnya setiap perkataan sinonim tersebut melalui proses yang sama iaitu perbandingan dengan index bagi setiap pantun. Jika terdapat perbandingan yang mengena tepat, pantun itulah yang akan

dipilih dan diberikan kepada pengguna.

Jika proses ini gagal juga mencari *index* yang tepat dengan kunci itu, maka proses seterusnya akan dilalui. Proses tersebut adalah proses expansion. Perkataan tambahan akan dikeluarkan oleh sistem sebagai cadangan atau suggestion kepada pengguna. Kemudian pengguna itu perlu memilih adakan expansion itu mengena dengan maksud yang pengguna inginkan. Apabila pengguna terima, maka proses pencarian pantun itu berjalan dengan perbandingan index. Kemudian pantunnya akan diberikan kepada pengguna tersebut.

Akhirnya jika sistem ini masih gagal untuk mencari pantun yang pengguna kehendaki, maka sistem ini akan respon bahawa pantun yang pengguna inginkan tiada dan belum lagi wujud di dalam database pantun ini.

## **HASIL KAJIAN**

### **4.1 PENGENALAN**

Pembangunan sistem pencarian pantun ini melibatkan pengumpulan data dalam bentuk data dan rangkap pantun yang setiapnya disimpan di dalam bentuk indek. Paparan yang akan digunakan untuk pembangunan sistem ini adalah menggunakan web agar ianya mudah diakses oleh pengguna di dalam setiap peranti dengan mudah. Peringkat implementasi dan penilaian enjin pencarian ini merupakan peringkat yang penting dalam memastikan sistem yang dibina berjalan dengan lancar dan efisien. Bab ini menjelaskan implementasi dan pengujian sistem ini.

### **4.2 HASIL IMPLEMENTASI**

Projek ini berhasrat untuk membangunkan teknik atau aplikasi yang boleh meramal pantun yang sesuai untuk kueri yang dimasukkan oleh pengguna. Pengguna dapat menggunakan sistem ini dengan paparan atau antara muka berasaskan web. Bagi pembangunan web, salah satu antara muka yang mudah untuk pengguna gunakan sistem ini. Antara muka web memudahkan pengguna dengan hanya mempunyai akses internet bagi membenarkan pengguna menggunakan sistem pencarian pantun ini. Maka, antara mukanya

adalah antara elemen yang penting dan perlu diambil kira bagi memastikan interaksi pengguna dengan sistem tinggi. Elemen penting yang terlibat dalam menampilkan antara muka yang menarik antaranya ialah penggunaan warna, font tulisan dan saiz tulisan, butang yang menonjol dan mudah untuk dilihat pengguna, serta hasil pencarian sistem ini juga perlu bersesuaian dengan pengguna. Rajah 4.1 menunjukkan antara muka bagi menu utama sistem ini.



Rajah 4.1 di atas menunjukkan menu utama sistem pencarian pantun.

Sistem ini menampilkan paparan yang simplistik dan minima. Penggunaan warnanya yang basic dan tidak terlalu banyak agar menampakkan konsistensi web ini. Font dan size tulisannya yang jelas dan mudah untuk dibaca main peranan penting dalam membantu interaksi pengguna dan sistem. Butang untuk ke paparan seterusnya jelas dan mudah untuk dicari.

#### 4.2.1 FRONTEND

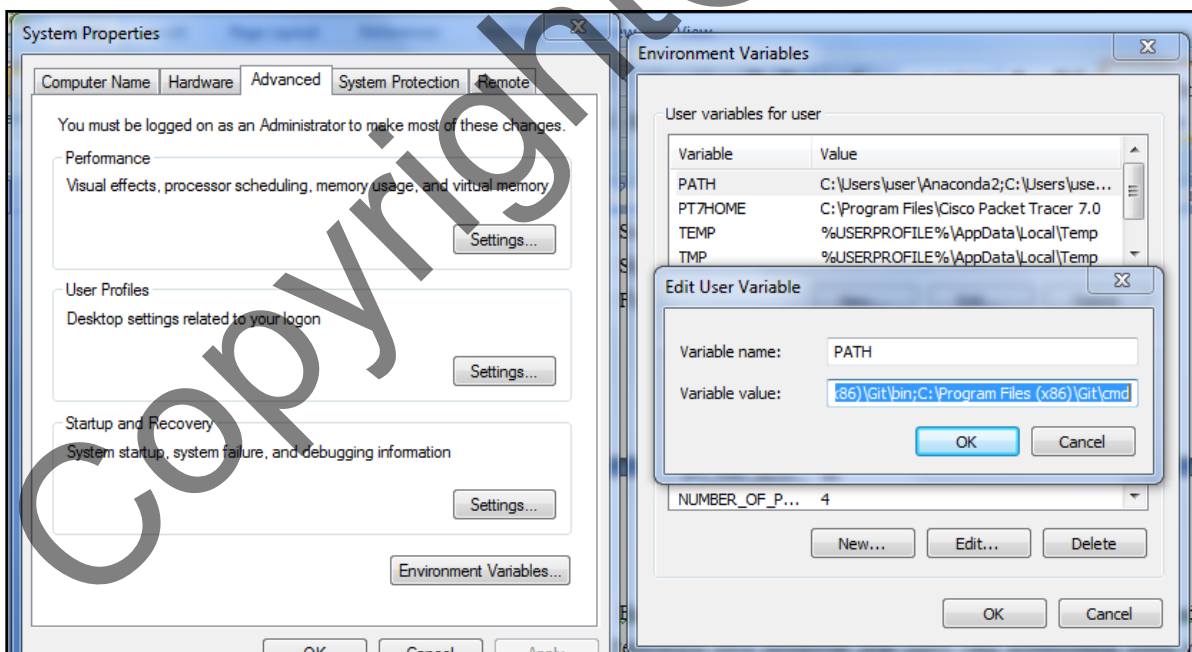
Bahagian FrontEnd merupakan peranti atau program yang diakses secara langsung oleh pengguna dan membolehkan pengguna mengakses ke peranti dan program yang seterusnya. Sistem ini dibangunkan dengan menggunakan *Anaconda2/Python2.7* di sini disertakan beberapa langkah yang perlu dijalankan sebelum membina paparan antara muka web.

Pembangun perlu untuk memuat turun beberapa modul bagi memanggil fungsi dan kod algoritma bagi membangun sistem ini. Antara modul yang diperlukan pembangun ialah:

- a) Pip (Python Installer Package)
- b) Sublime
- c) SQLite
- d) Flask-PonyWhoosh

Bagi memastikan sistem dapat dibangunkan, sistem perlu berapa pada jalan (*path*) yang betul. Pembangun perlu menambah jalan (*path*) yang membolehkan sistem membaca *Python* dan *Anaconda*. Cara untuk menambah adalah dengan membuka pada bahagian panel kawalan (*control panel*). Seterusnya, pergi ke ruangan **Control Panel>System and Security>System**. Di sana terdapat butang **Advance System Setting>Enviroment Variables..>Path**. Pembangun perlu untuk menambah perkara dibawah bagi membenarkan sistem menjalankan *Anaconda2* dan *Python2.7*.

(C:\Users\user\Anaconda2;C:\Users\user\Anaconda2\Scripts;C:\Users\user\Anaconda2\Library\bin;C:\Program Files (x86)\Git\bin;C:\Program Files (x86)\Git\cmd)



Rajah 4.2 Panduan menukar dan menambah direktori pada jalan (*path*)

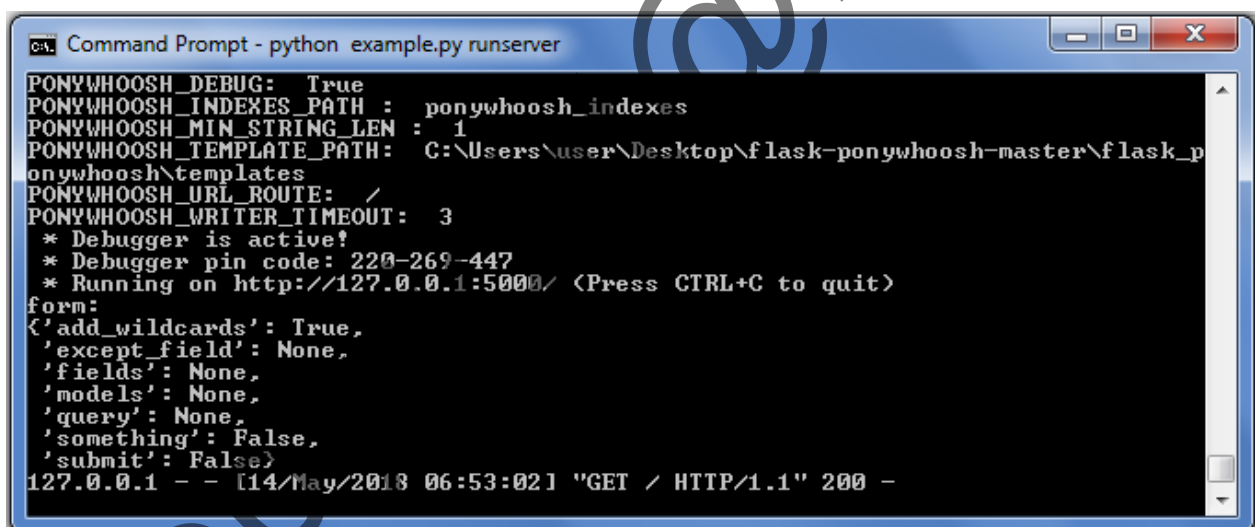
Setelah selesai, pembangun perlu meletakkan modul dan fail yang telah disediakan pada direktori yang pembangun inginkan dan memudahkan untuk dijalankan. Seterusnya

pembangun perlu membuka *Command Prompt* untuk memilih modul yang perlu dijalankan. Masukkan kod di bawah dengan direktori yang betul :

- cd C:\Users\user\Desktop\flask-ponywhoosh-master
- pip install -r requirements.txt
- python example.py runserver

Seterusnya, pembangun perlu menjalankan membuka *localhost* dan pautan *web sistem* ini. Buka *web browser* dan buka link <http://127.0.0.1:5000/>. Setelah link itu dibuka maka teraparlah paparan antara muka sistem ini berdasarkan pembangunan web.

Rajah 4.4 adalah paparan *Command prompt* setelah program selesai berjalan maka sistem boleh serta merta digunakan. Seterusnya pengguna boleh terus memasukkan kueri pada link itu dan menjalankan pencarian pantun mengikut kehendak pengguna.



```

C:\> Command Prompt - python example.py runserver
PONYWHOOSH_DEBUG: True
PONYWHOOSH_INDEXES_PATH : ponywhoosh_indexes
PONYWHOOSH_MIN_STRING_LEN : 1
PONYWHOOSH_TEMPLATE_PATH: C:\Users\user\Desktop\flask-ponywhoosh-master\flask_p
onywhoosh\templates
PONYWHOOSH_URL_ROUTE: /
PONYWHOOSH_WRITER_TIMEOUT: 3
* Debugger is active!
* Debugger pin code: 220-269-447
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
form:
{'add_wildcards': True,
 'except_field': None,
 'fields': None,
 'models': None,
 'query': None,
 'something': False,
 'submit': False}
127.0.0.1 - - [14/May/2018 06:53:02] "GET / HTTP/1.1" 200 -

```

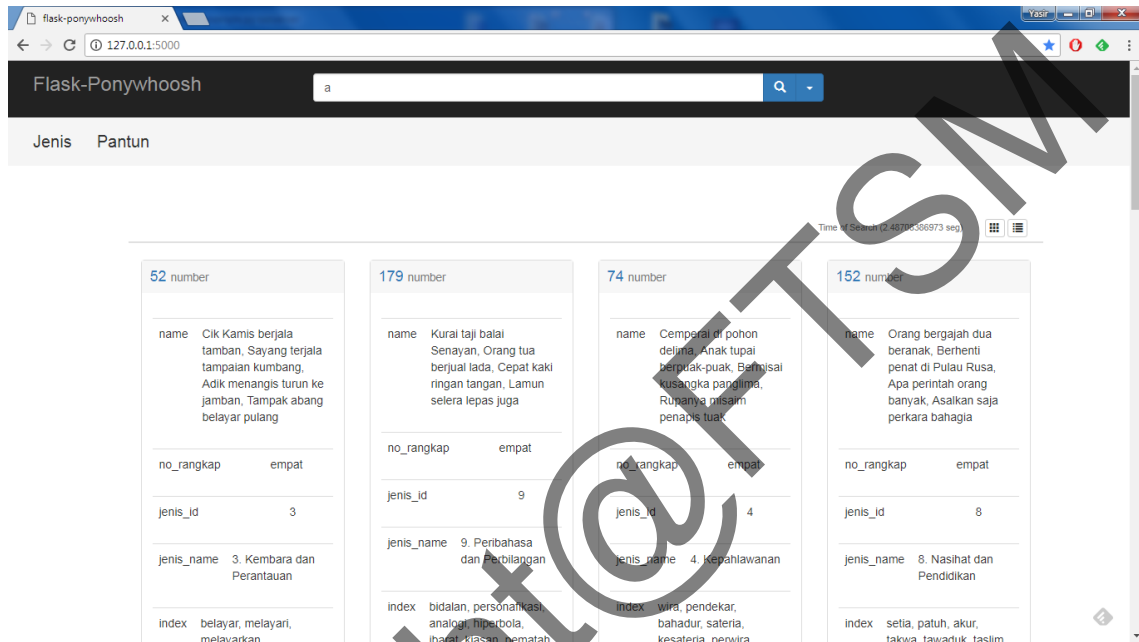
RAJAH 4.3 paparan *Command prompt* setelah program selesai berjalan

Cara enjin gelintar ini berfungsi adalah bergantung sepenuhnya kepada kueri yang dimasukkan oleh pengguna. Untuk bahagian carian, pengguna hanya perlu memasukkan apa-apa jenis kueri termasuk karakter secara tunggal atau dalam satu perkataan lengkap. Sistem ini juga dapat menghasilkan output daripada kueri dalam bentuk ayat penuh. Kueri yang padan dan berkaitan dengan pantun akan mengeluarkan hasilnya (output) senarai 10 pantun yang berkaitan dengan kueri. Kueri yang dapat dikesan dan dapat dikenalpasti oleh sistem ini adalah kueri dalam bentuk karakter tunggal (nombor dan huruf selain simbol) kueri dalam



bentuk satu perkataan penuh dan kueri yang mempunyai dua atau lebih perkataan yang dimasukkan oleh pengguna. Sistem ini juga membaca kueri bagi setiap terpisahnya karakter atau perkataan yang dipisahkan dengan karakter jarak (*spacebar*).

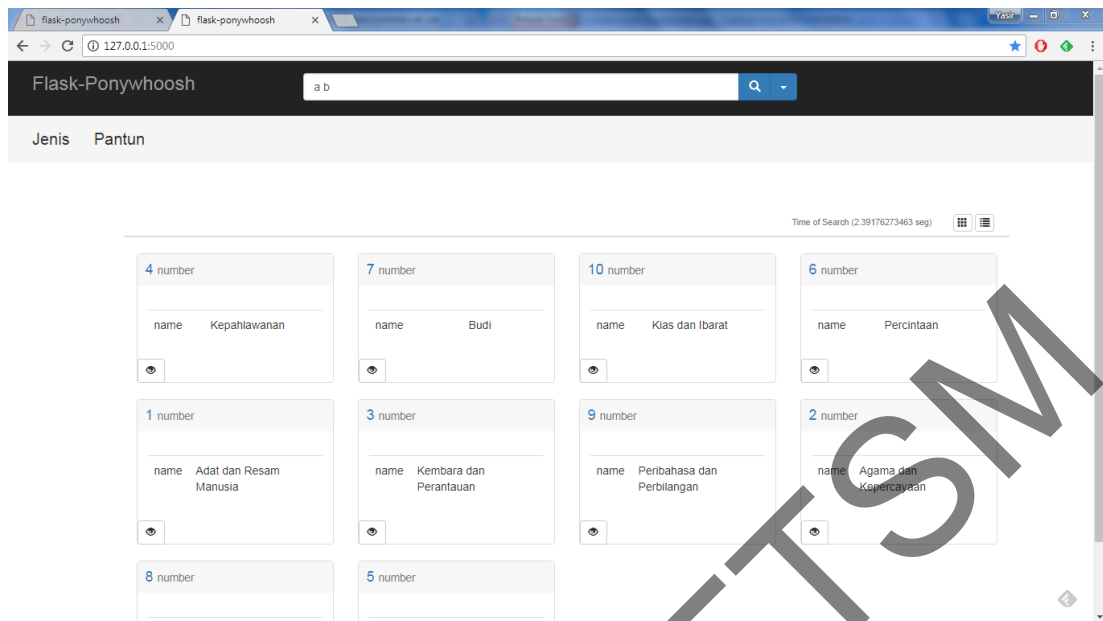
#### A) Kueri dalam bentuk karakter tunggal.



RAJAH 4.4 Antara muka menunjukkan kueri satu karakter tunggal yang dimasukkan oleh pengguna

Pantun yang mempunyai karakter huruf 'a' yang paling banyak akan berada pada baris yang pertama mengikut turutan dan kedudukan (*ranking*) yang lebih sekaligus memberi hasil kepada kueri yang dimasukkan oleh pengguna. Menggunakan sistem ranking, pantun yang mempunyai karakter 'a' paling banyak keluar dalam ayat akan berada di ranking yang teratas, dan diikuti dengan pantun lain ke bawah berdasarkan kepada kekerapan (frequency) perkataan yang dimasukkan oleh pengguna (kueri) keluar dalam pantun. Sistem ini mencampurkan kesemua huruf 'a' yang wujud termasuk indeks yang berada di dalam nama pantun, nombor rangkap pantun, id jenis pantun, nama jenis pantun dan indeks tambahan. Karakter nombor juga boleh dicari oleh sistem kecuali karakter simbol.

## B) Kueri dalam bentuk karakter tunggal dua atau lebih.



RAJAH 4.5 Antara muka menunjukkan kueri satu karakter tunggal yang dimasukkan oleh pengguna

Sistem pencarian pantun ini ia dapat mencari berbilang karakter kueri yang pengguna masukkan sebagai kueri-kueri yang tunggal. Contoh yang diberikan ialah pencarian karakter berbilang iaitu 'a b'. Sistem ini mencari pantun yang berkaitan dengan mencari karakter yang padan dengan kueri 'a', 'b' dan 'a b'. Malah, cara pantun itu disusun dengan kedudukan (*ranking*) tidaklah diletakkan pantun yang mempunyai kedua-dua perkataan tersebut sebagai kueri yang tertinggi, tetapi pantun yang dikira kebarangkaliannya besar merupakan pantun yang padan dipilih. Perkara yang sama juga berlaku iaitu sistem ini mencampurkan kesemua perkataan 'a', 'b' dan 'a b' yang wujud termasuk indeks yang berada di dalam nama pantun, nombor rangkap pantun, id jenis pantun, nama jenis pantun dan indeks tambahan.

## C) Kueri dalam bentuk satu perkataan lengkap

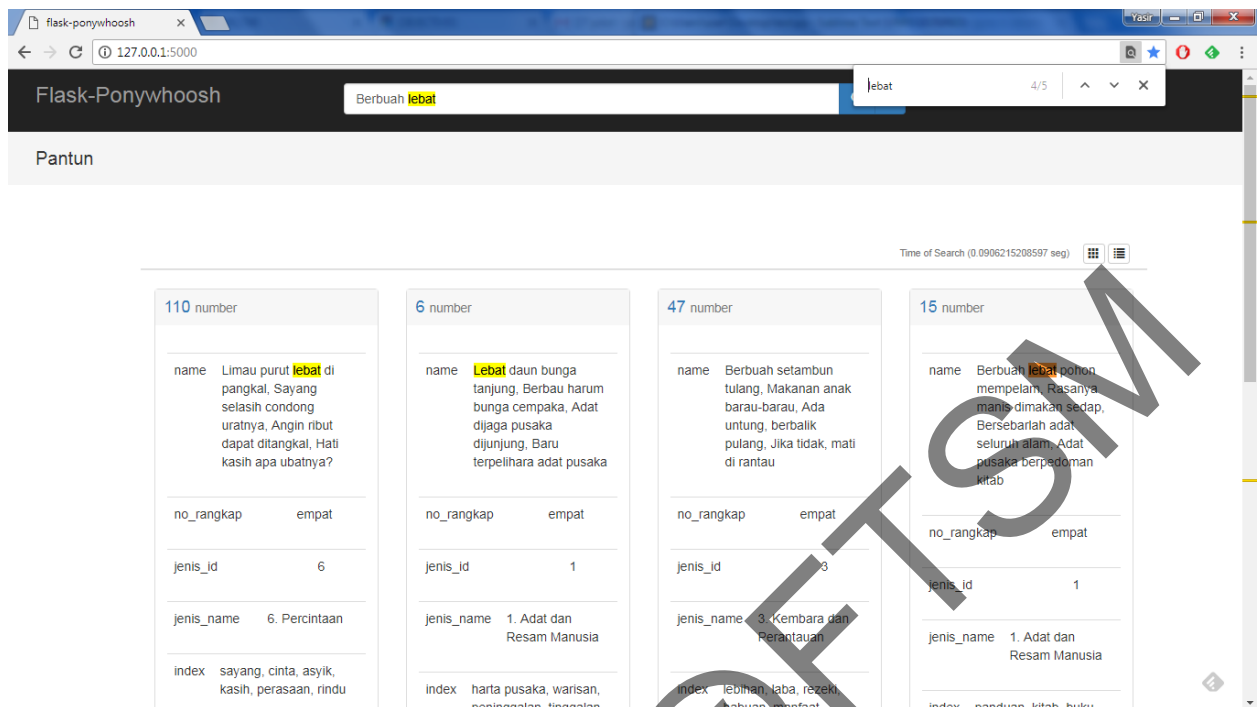
The screenshot shows a web browser window with the URL '127.0.0.1:5000' and the page title 'Flask-Ponywhoosh'. The search bar contains the query 'baik'. Below the search bar, the word 'Pantun' is displayed. The search results are presented in a grid of four cards, each representing a different pantun entry. Each card includes a count of occurrences, a full text of the pantun, and various metadata fields such as 'no\_rangkap', 'jenis\_id', 'jenis\_name', and 'index'.

Count	Name	no_rangkap	jenis_id	jenis_name	index
72	Baik-baik mengail tenggiri, Takut terkena ikan parang, Baik-baik merendah diri, Jangan menjadi hamba orang	empat	4	4. Kepahlawanan	lemah lembut, budi pekerti, baik hati
13	Gerang perat pada sutera, Pandan sehelai di dalam jerami, Baik paras Inderaputera, Baik diambil akan suami	empat	1	1. Adat dan Resam Manusia	anak raja lelaki, anak lelaki
161	Tuan puteri duduk berteduh, Di pinggir kota naungan pupur, Lepas mencari bunga keladi, Suami isteri kalau bergaduh, Usah kita turut campur, Kita sendiri kedapatan budi	enam	8	8. Nasihat dan Pendidikan	
12	Baik-baik mengail tenggiri, Jangan kena jambatan tarangan, Baik-baik menghina diri, Jangan muka diapu arang	empat	1	1. Adat dan Resam Manusia	keji, tercela, tercemar, terkutuk, lata, buruk

RAJAH 4.6 Antara muka menunjukkan kueri satu perkataan yang dimasukkan oleh pengguna

Pantun yang mempunyai perkataan 'baik' yang paling banyak akan berada pada baris yang pertama mengikut turutan dan kedudukan (*ranking*) yang lebih sekaligus memberi hasil kepada kueri yang dimasukkan oleh pengguna. Menggunakan sistem ranking, pantun yang mempunyai perkataan 'baik' paling banyak keluar dalam ayat akan berada di ranking yang teratas, dan diikuti dengan pantun lain ke bawah berdasarkan kepada kekerapan (frequency) perkataan yang dimasukkan oleh pengguna (kueri) keluar dalam pantun. Sistem ini mencampurkan kesemua perkataan 'baik' yang wujud termasuk indeks yang berada di dalam nama pantun, nombor rangkap pantun, id jenis pantun, nama jenis pantun dan indeks tambahan.

## D) Kueri dalam bentuk dua perkataan atau lebih



RAJAH 4.7 Antara muka menunjukkan kueri dua atau lebih perkataan yang dimasukkan oleh pengguna

Sistem pencarian pantun ini ia dapat mencari berbilang perkataan kueri yang pengguna masukkan sebagai kueri-kueri yang tunggal. Contoh yang diberikan ialah pencarian perkataan berbilang iaitu 'berbuah lebat'. Sistem ini mencari pantun yang berkaitan dengan mencari perkataan yang padan dengan kueri 'berbuah', 'lebat' dan 'berbuah lebat'. Malah, cara pantun itu disusun dengan kedudukan (*ranking*) tidaklah diletakkan pantun yang mempunyai kedua-dua perkataan tersebut sebagai kueri yang tertinggi, tetapi pantun yang dikira kebarangkaliannya besar merupakan pantun yang padan dipilih. Maka kekurangan di sini ialah pantun yang terdapat kedua-dua perkataan yang kene dengan kehendak pengguna tidak dipilih. Perkara yang sama juga berlaku iaitu sistem ini mencampurkan kesemua perkataan 'berbuah', 'lebat' dan 'berbuah lebat' yang wujud termasuk indeks yang berada di dalam nama pantun, nombor rangkap pantun, id jenis pantun, nama jenis pantun dan indeks tambahan.

## 4.2.2 KOD GELINTARAN

Backend merupakan pemrosesan atau program bawahan secara tidak langsung yang diakses oleh pengguna. Bahagian ini melaksanakan fungsi khusus bagi pihak pemproses utama atau sistem perisian. Di sini tempat untuk membuat reka bentuk paparan web yang akan dipaparkan sebagai sistem kepada pengguna atau dipanggil *FrontEnd*.

```

1  {% extends "ponywhoosh/base.html" %}
2  {% import "bootstrap/wtf.html" as wtf %}
3  {% block styles %}
4  {{ super() }}
5  <style>
6  .container-search{
7      text-align: center;
8  }
9  h1{
10     font-size: 600%;
11     font-family: 'Bree Serif', serif;
12 }
13 .btn-avan{
14     width: 100px;
15     margin-left: 45%;
16 }
17
18 #buscador{
19     margin-left: 510px;
20 }
21 @media screen and (max-width: 768px) {
22     .title-search {
23         font-size: 500%;
24     }
25     .btn-avan {
26         margin-left: 30%;
27     }
28 }
29 </style>
30 {% endblock %}
31 {% block title %}Pencarian Pantun | flask-ponywhoosh{% endblock %}
32 {% block page_content %}
33 <link href="https://fonts.googleapis.com/css?family=Bree+Serif"
34     rel="stylesheet"
35     type="text/css">
36 <div class="page-header" style="margin-top: 70px">
37     <div class="container container-search">
38         <div class="row">
39             <div class="col-md-12 col-xs-12">
40                 <h1 class="title-search">
41                     Pantun <small><h4>Selamat datang ke sistem pencarian pantun</h4></small>
42                 </h1>
43             </div>
44         </div>
45     </div>
46 </div>
47 <div class="container">
48     <form method="POST" action={{action_url_form}} role="search" id="form-search">
49         {{ form.hidden_tag() }}
50     </form>
51     <div class="row">
52

```

RAJAH 4.8 Kod utama untuk pembangunan web.

Di sini juga terlibat juga proses penambahan indeks dan database senarai pantun beserta segala skema pantun yang diperlukan antaranya nama pantun, nombor jenis pantun, nama jenis pantun dan indeks. Pantun yang telah tersedia di dalam sistem ini adalah sejumlah 200 pantun yang mempunyai 10 jenis atau kategori.

```

88 if select(s for s in Pantun).count() > 8: return
89 d1 = Jenis(name="Adat dan Resam Manusia")
90 d2 = Jenis(name="Agama dan Kepercayaan")
91 d3 = Jenis(name="Kembara dan Perantauan")
92 d4 = Jenis(name="Kepahlawanan")
93 d5 = Jenis(name="Teku-teki")
94 d6 = Jenis(name="Percintaan")
95 d7 = Jenis(name="Budai")
96 d8 = Jenis(name="Nasihat dan Pendidikan")
97 d9 = Jenis(name="Peribahasa dan Perbilangn")
98 dx = Jenis(name="kias dan Ibarat")
99
100
101
102 # 1. Adat dan Resam Manusia
103 # dua
104 p101 = Pantun(
105     name="Taptibau puyuh padang, Hilang pisau berganti parang",
106     jenis_id=d1,
107     jenis_name="j1",
108     no_rangkap="dua",
109     indeks="badik, belati, belintan, cundrik, curik, kerambit, lawi ayam, nenraa, pisau calung, pisau cukur, pisau getah, pisau lipat, pisau makan, pisau ra"
110 )
111
112
113 p102 = Pantun(
114     name="Banyak udang banyak garam, Banyak orang banyak ragam",
115     jenis_id=d1,
116     jenis_name="j1",
117     no_rangkap="dua",
118     indeks="olah: tingkah, kerenah, perandai, gelagat, canda, fiil,"
119 )
120
121
122 p103 = Pantun(
123     name="Pagi petang siang malam, Hati terang senang faham",
124     jenis_id=d1,
125     jenis_name="j1",
126     no_rangkap="dua",
127     indeks="perasaan, jiwa, roh, emosi"
128 )
129
130 # empat
131 p104 = Pantun(
132     name="Langkah bulat batang meranti, Nke dalam air tampak bayangan, nBerkat daulat bersisi sakti, nKasalnya turun dari kayangan.",
133     jenis_id=d1,
134     jenis_name="j1",
135     no_rangkap="dua",
136     indeks="keramat, bertuah, mujarab, manjur, berkuasa, berpuaka, ampuh"
137 )
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2593
2594
2595
2596
2597
259
```

### 4.3 HASIL PENGUJIAN

Hasil pengujian adalah dokumen yang memperincikan objektif, sasaran, paparan dan proses untuk enjin gelintar tersebut. Pelan pengujian diperlukan untuk melaksanakan pengujian.

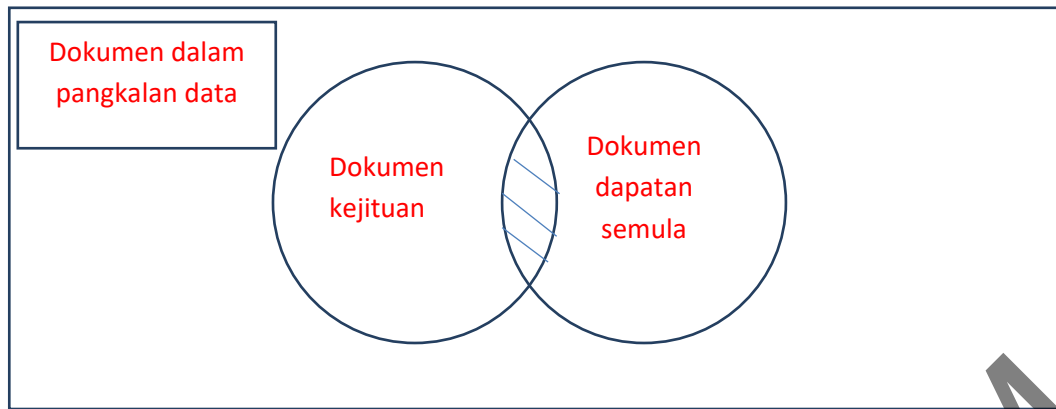
#### 4.3.1 Tujuan Pengujian

- i. Mengenal pasti kekurangan projek dan penambahbaikan yang boleh dibuat
- ii. Memastikan enjin gelintar yang dibangunkan dapat memenuhi objektif kajian
- iii. Fungsi yang terdapat dalam sistem berjalan dengan lancar
- iv. Untuk melihat sekiranya projek dapat memberi manfaat kepada pengguna

Bagi mengukur sesuatu sistem capaian maklumat (*Information Retrieval System*), beberapa langkah perlu diambil bagi memastikan kejituan (*precision*) dan dapatan semula (*recall*) sesebuah enjin gelintar berada pada tahap yang seimbang. Penyediaan koleksi ujian (*Test Collections*) perlu dilakukan terlebih dahulu. Koleksi ini terdiri daripada set dokumen, kueri, dan juga penilaian yang berkaitan. Hal ini dapat membenarkan pembangun membuat eksperimen secara berulang atau eksperimen secara perbandingan.

#### 4.3.2 Kaedah Pengujian

Setiap dokumen yang tersimpan dalam pangkalan data boleh dipanggil berdasarkan kueri oleh pengguna. Output yang terhasil adalah terbahagi kepada tiga, iaitu dokumen relevan (*relevant documents*), dokumen yang dicapai (*retrieved documents*), serta dokumen yang dicapai & relevan (*relevant & retrieved documents*).



RAJAH 4.10 Rajah Klasifikasi Dokumen Yang Dicapai

Untuk menjalankan kaedah kajian yang dibincangkan, dua enjin gelintar telah dibangunkan bagi membuktikan kemampuan sistem yang diubah suai mengatasi sistem yang sedia ada. Sistem enjin gelintar yang telah diubah suai ( $E_2$ ) mengandungi algoritma untuk membantu pengguna mencapai maklumat mengenai perkataan yang seakan ataupun hampir sama (sinonim) berbanding dengan enjin carian biasa ( $E_1$ ) yang hanya boleh mencapai perkataan yang 100% sama dengan kueri yang dimasukkan oleh pengguna. Enjin gelintar ini mengaplikasikan perkembangan indeks (*index expansion*) di mana sinonim diperkenalkan dalam proses pengindeksan.

Formula bagi kejitaan dan dapatan semula (*precision and recall*) adalah seperti berikut:

$$\text{Dapatan Semula} = \frac{\text{Bilangan dokumen relevan yang diambil semula}}{\text{Jumlah dokumen relevan yang wujud}}$$

$$\text{Kejitaan} = \frac{\text{Bilangan dokumen relevan yang diambil semula}}{\text{Jumlah dokumen yang diambil semula}}$$

Untuk mempertimbangkan koleksi rujukan dan set pengujian kueri, perkataan 'baik', 'adat', dan 'pendekar' akan mewakili  $K_1$ ,  $K_2$  dan  $K_3$  sebagai kueri yang akan digunakan dalam satu set kueri untuk proses ujian (*testing*).



Ketiga-tiga perkataan ini boleh membawa maksud yang berbeza atau dalam erti kata lain, masing-masing mempunyai pengertian yang berbeza. Akan tetapi, untuk proses ujian ini, perkataan '**baik**' akan membawa erti kepada '**untung, faedah, lebihan, rezeki, habuan, mujur, manfaat**', perkataan '**adat**' membawa maksud '**kelaziman, istiadat, amalan, adab, tradisi, upacara,**', manakala '**pahlawan**' membawa maksud '**wira, sateria, kesateria, perwira, panglima, perajurit, pejuang,**'. Berikut merupakan set dokumen yang relevan mengikut set pengujian kueri:

Relavan sebagai R,

$R_{K_1}$  sebagai dokumen yang relevan bagi kueri 1

$R_{K_2}$  sebagai dokumen yang relevan bagi kueri 2

$R_{K_3}$  sebagai dokumen yang relevan bagi kueri 3

$R_{K_1} : \{d_9, d_{47}, d_{48}, d_{72}, d_{131}, d_{132}, d_{133},\}$

$R_{K_2} : \{d_6, d_7, d_8, d_{10}, d_{15}, d_{16}, d_{149},\}$

$R_{K_3} : \{d_{70}, d_{74}, d_{77}, d_{81},\}$

Dokumen-dokumen yang dicapai akan dihadkan kepada 10 dokumen sahaja bagi memnudahkan pengiraan. Sebarang persamaan yang terdapat pada output bagi enjin gelintar  $E_1$  dan  $E_2$  bagi  $K_1$ ,  $K_2$  dan  $K_3$  dengan  $R_{K_1}$ ,  $R_{K_2}$  dan  $R_{K_3}$ , masing-masing akan ditandakan dengan bulatan sebagai dokumen yang relevan mengikut ujian pertama, kedua dan ketiga.

### Ujian Pertama (kueri: '**baik**')

Enjin 1 :  $\{d_9, d_{127}, d_{13}, d_{145}, d_{12}, d_{69}, d_{122}, d_{128}, d_{133}, d_{129}\}$

Enjin 2 :  $\{d_{72}, d_{13}, d_{16}, d_{12}, d_{47}, d_{133}, d_{48}, d_{132}, d_{134}, d_9\}$

$$R_{K_1} : \{d_9, d_{47}, d_{48}, d_{72}, d_{131}, d_{132}, d_{133}\}$$

Seterusnya, dokumen yang dianggap relevan dan dapat dicapai (dokumen yang ditebalkan) akan diukur mengikut urutan *ranking* di mana dokumen yang paling awal dicapai akan dikira terlebih dahulu. Setiap hasil akan direkodkan. Jadual di bawah menunjukkan hasil kiraan:

JADUAL 4.1 Jadual Perbandingan  $E_1$  dengan  $E_2$  Berdasarkan Kejituan Dokumen Yang Dicapai Bagi  $K_1$

Dokumen relevan (enjin 1) :  $\{d_9, d_{133}\}$

Kedudukan	Enjin 1 ( $E_1$ )		Enjin 2 ( $E_2$ )	
	Kejituan	Dapatan Semula	Kejituan	Dapatan Semula
1	<b>1/1 = 1.00</b>	<b>1/6 = 0.17</b>	<b>1/1 = 1.00</b>	<b>1/6 = 0.17</b>
2	0/2 = 0.00	1/6 = 0.17	0/2 = 0.00	1/6 = 0.17
3	0/3 = 0.00	1/6 = 0.17	0/3 = 0.00	1/6 = 0.17
4	0/4 = 0.00	1/6 = 0.17	0/4 = 0.00	1/6 = 0.17
5	0/5 = 0.00	1/6 = 0.17	<b>2/5 = 0.40</b>	<b>2/6 = 0.33</b>
6	0/6 = 0.00	1/6 = 0.17	<b>3/6 = 0.50</b>	<b>3/6 = 0.50</b>
7	0/7 = 0.00	1/6 = 0.17	<b>4/7 = 0.57</b>	<b>4/6 = 0.67</b>
8	0/8 = 0.00	1/6 = 0.17	<b>5/8 = 0.63</b>	<b>5/6 = 0.83</b>
9	<b>2/9 = 0.22</b>	<b>2/6 = 0.33</b>	0/9 = 0.00	5/6 = 0.83
10	0/10 = 0.00	2/6 = 0.33	<b>7/10 = 0.70</b>	<b>6/6 = 1.00</b>

$$\text{Kejituan purata} = \text{Ranking (enjin 1)} = \frac{1.00 + 0.20}{2} = 0.60$$

Dokumen relevan (enjin 2) :  $\{d_{72}, d_{47}, d_{133}, d_{48}, d_{132}, d_9\}$

$$\text{Kejituan purata} = \text{Ranking (enjin 2)} = \frac{1.00 + 0.40 + 0.50 + 0.57 + 0.63 + 0.70}{6} = 0.63$$

**Ujian Kedua (kueri: ‘adat’)**

Enjin 1 :  $\{d_{146}, d_{167}, \mathbf{d_7}, \mathbf{d_6}, d_{46}, d_{28}, \mathbf{d_{16}}, d_{123}, \mathbf{d_8}, d_5\}$

Enjin 2 :  $\{\mathbf{d_{149}}, \mathbf{d_7}, \mathbf{d_{10}}, \mathbf{d_6}, d_{27}, d_{18}, \mathbf{d_{16}}, d_{17}, d_{49}, \mathbf{d_8}\}$

$R_{K_2} : \{\mathbf{d_6}, \mathbf{d_7}, \mathbf{d_8}, \mathbf{d_{10}}, \mathbf{d_{15}}, \mathbf{d_{16}}, \mathbf{d_{149}},\}$

Seterusnya, dokumen yang dianggap relevan dan dapat dicapai (dokumen yang ditebalkan) akan diukur mengikut urutan *ranking* di mana dokumen yang paling awal dicapai akan dikira terlebih dahulu. Setiap hasil akan direkodkan. Jadual di bawah menunjukkan hasil kiraan:

JADUAL 4.2 Jadual Perbandingan  $E_1$  dengan  $E_2$  Berdasarkan Kejituan Dokumen Yang Dicapai Bagi  $K_1$

Dokumen relevan (enjin 1) :  $\{\mathbf{d_7}, \mathbf{d_6}, \mathbf{d_{16}}, \mathbf{d_8},\}$

Kedudukan	Enjin 1 ( $E_1$ )		Enjin 2 ( $E_2$ )	
	Kejituan	Dapatan Semula	Kejituan	Dapatan Semula
1	0/1 = 0.00	0/7 = 0.00	<b>1/1 = 1.00</b>	<b>1/7 = 0.14</b>
2	0/2 = 0.00	0/7 = 0.00	<b>2/2 = 1.00</b>	<b>2/7 = 0.29</b>
3	<b>1/3 = 0.33</b>	<b>1/7 = 0.14</b>	<b>3/3 = 1.00</b>	<b>3/7 = 0.43</b>
4	<b>2/4 = 0.50</b>	<b>2/7 = 0.29</b>	<b>4/4 = 1.00</b>	<b>4/7 = 0.57</b>
5	0/5 = 0.00	2/7 = 0.29	0/5 = 0.00	4/7 = 0.57
6	0/6 = 0.00	2/7 = 0.29	0/6 = 0.00	4/7 = 0.57
7	<b>3/7 = 0.43</b>	<b>3/7 = 0.43</b>	<b>5/7 = 0.71</b>	<b>5/7 = 0.71</b>
8	0/8 = 0.00	3/7 = 0.43	0/8 = 0.00	5/7 = 0.71
9	<b>4/9 = 0.44</b>	<b>4/7 = 0.57</b>	0/9 = 0.00	5/7 = 0.71
10	0/10 = 0.00	4/7 = 0.57	<b>6/10 = 0.60</b>	<b>6/7 = 0.86</b>

Kejituan purata = *Ranking* (enjin 1) =  $\frac{0.33+0.50+0.43+0.44}{4} = 0.425$

Dokumen relevan (enjin 2) :  $\{\mathbf{d_{149}}, \mathbf{d_7}, \mathbf{d_{10}}, \mathbf{d_6}, \mathbf{d_{16}}, \mathbf{d_8}\}$

$$\text{Kejituan purata} = \text{Ranking (enjin 2)} = \frac{1.00 + 1.00 + 1.00 + 1.00 + 0.71 + 0.60}{6} = 0.885$$

### Ujian Ketiga (kueri: ‘pahlawan’)

Enjin 1 : {  $d_{88}$ ,  $d_{81}$ ,  $d_{67}$ ,  $d_{79}$ ,  $d_{78}$ ,  $d_{77}$ ,  $d_{76}$ ,  $d_{75}$ ,  $d_{74}$ ,  $d_{73}$  }

Enjin 2 : {  $d_{74}$ ,  $d_{81}$ ,  $d_{70}$ ,  $d_{84}$ ,  $d_{79}$ ,  $d_{78}$ ,  $d_{77}$ ,  $d_{83}$ ,  $d_{73}$ ,  $d_{81}$  }

$R_{K_3}$  : {  $d_{70}$ ,  $d_{74}$ ,  $d_{77}$ ,  $d_{81}$ , } }

Seterusnya, dokumen yang dianggap relevan dan dapat dicapai (dokumen yang ditebalkan) akan diukur mengikut urutan *ranking* di mana dokumen yang paling awal dicapai akan dikira terlebih dahulu. Setiap hasil akan direkodkan. Jadual menunjukkan hasil kiraan:

JADUAL 4.3 Jadual Perbandingan  $E_1$  dengan  $E_2$  Berdasarkan Kejituan Dokumen Yang Dicapai Bagi  $K_1$

Dokumen relevan (enjin 1) : {  $d_{81}$ ,  $d_{77}$ ,  $d_{74}$  }

$$\text{Kejituan purata} = \text{Ranking (enjin 1)} = \frac{0.50 + 0.33 + 0.33}{3} = 0.39$$

Kedudukan	Enjin 1 ( $E_1$ )		Enjin 2 ( $E_2$ )	
	Kejituan	Dapatan Semula	Kejituan	Dapatan Semula
1	0/1 = 0.00	0/4 = 0.00	<b>1/1 = 1.00</b>	<b>1/4 = 0.25</b>
2	<b>1/2 = 0.50</b>	<b>1/4 = 0.25</b>	<b>2/2 = 1.00</b>	<b>2/4 = 0.50</b>
3	0/3 = 0.00	1/4 = 0.25	0/3 = 0.00	2/4 = 0.50
4	0/4 = 0.00	1/4 = 0.25	0/4 = 0.00	2/4 = 0.50
5	0/5 = 0.00	1/4 = 0.25	0/5 = 0.00	2/4 = 0.50
6	<b>2/6 = 0.33</b>	<b>2/4 = 0.50</b>	0/6 = 0.00	2/4 = 0.50
7	0/7 = 0.00	2/4 = 0.50	<b>3/7 = 0.43</b>	<b>3/4 = 0.75</b>
8	0/8 = 0.00	2/4 = 0.50	0/8 = 0.00	3/4 = 0.75
9	<b>3/9 = 0.33</b>	<b>3/4 = 0.75</b>	0/9 = 0.00	3/4 = 0.75
10	0/10 = 0.00	3/4 = 0.75	<b>4/10 = 0.40</b>	<b>4/4 = 1.00</b>

Dokumen relevan (enjin 2) : {  $d_{74}$ ,  $d_{81}$ ,  $d_{77}$ ,  $d_{81}$  }

$$\text{Kejituan purata} = \text{Ranking (enjin 2)} = \frac{1.00+1.00+0.43+0.40}{4} = 0.71$$

Hasil daripada pengiraan dan pengujian ini menampakkan bahawa enjin pencarian yang ditambah dengan *feature* tambahan iaitu index expansion mengeluarkan hasil yang lebih hamper dan mendapat markah bacaan yang tinggi dan tepat dengan ketepatan sebenar. Walaupun enjin yang ditambahbaik ini berhasil, tetapi masih juga tidak dapat untuk mengeluarkan semua pantun yang berkaitan dan terdapat juga pantun yang tidak dikategorikan sebagai mirip juga turut dikeluarkan.

#### 4.4 KESIMPULAN

Keseluruhan bab ini menerangkan secara terperinci mengenai pengujian untuk pembangunan enjin gelintar ini berdasarkan pembangunan web. Antara muka modul juga turut dipaparkan. Selain itu, proses pengujian ini dijalankan untuk mengenal pasti kemampuan enjin gelintar yang dibina mengatasi enjin carian biasa dari segi capaian dokumen yang relevan. Akhir sekali, fasa pengujian juga dapat memastikan pembangunan sistem dapat dibangunkan dan berfungsi dengan baik.

#### 5.1 RINGKASAN PROJEK

Sistem pencarian pantun yang sesuai ini ingin dibangunkan dengan objektif untuk membantu golongan mahasiswa dalam menyiapkan tugas mereka berkaitan penggunaan pantun dengan menggunakan teknik tersendiri untuk meramal penggunaan pantun yang betul dengan maksud yang diinginkan.

Antara fungsi yang disediakan oleh sistem ini termasuklah fungsi memasukkan indeks, di mana fungsi ini memasukkan dokumen yang mengandungi segala data (pantun dan contoh ayat) dan disimpan dalam pangkalan data dalam format XML. Fungsi yang kedua ialah fungsi memasukkan kueri iaitu fungsi untuk pengguna memasukkan kata kunci, secara

tidak langsung mengakses indeks dalam pangkalan data dan memperoleh data (output) yang menepati kueri yang dimasukkan oleh pengguna.

Fungsi-fungsi lain yang wujud di sistem ini termasuklah fungsi log masuk dan fungsi cadangan.

## 5.2 KEKANGAN DALAM PROJEK

### 1. Fokus kepada satu kategori sahaja (pantun sahaja)

- Data yang dikumpul hanya difokuskan kepada pantun sahaja, dan bukannya pantun secara menyeluruh, yang secara umumnya mengandungi empat bahagian sahaja, iaitu pantun, bidalan, pepatah dan perumpamaan

### 2. Penggunaan contoh ayat dihadkan kepada lima sahaja

- Setiap pantun dihadkan jumlah ayat contohnya kepada lima ayat sahaja

### 3. Data dalam bentuk pantun (pantun) dihadkan kepada 100 sahaja

- Jumlah pantun yang dapat dikumpul secara keseluruhan adalah sebanyak 100 sahaja

### 4. Skop projek dihadkan kepada mahasiswa dalam UKM sahaja

- Pengguna yang ditujukan terdiri daripada mahasiswa dari fakulti FSSK & FTSM sahaja

## 5.3 PENYELESAIAN YANG DICADANGKAN

Projek ini berhasrat untuk membangunkan teknik atau aplikasi yang boleh meramal pantun yang sesuai untuk sesuatu ayat. Pelajar perlu mengumpul dan membangunkan set korpus (pangkalan data) pantun yang dikhususkan kepada simpulan bahasa dan contoh penggunaan ayat. Korpus ini akan digunakan oleh teknik yang bersesuaian sama ada teknik pembelajaran mesin, teknik capaian maklumat atau teknik ramalan dalam menentukan pantun yang bersesuaian mengikut ayat yang dimasukkan oleh pengguna. Dengan berbekalkan kata kunci yang sesuai yang terdapat dalam ayat

yang dimasukkan, sistem akan mengesan kata kunci tersebut, dan terus menghubungkan dengan set korpus, dan secara tidak langsung, mengesan pantun untuk ayat tersebut secara automatik. Ini juga secara tidak langsung memudahkan pengguna, khususnya mahasiswa dan mahasiswi di dalam Universiti Kebangsaan Malaysia untuk mengaplikasikan penggunaan sistem ini dalam urusan menyiapkan tugas mereka (pelajar aliran Bahasa khususnya).

## **KESIMPULAN**

Sistem Mengesan Pantun Yang Sesuai ini dapat membantu golongan mahasiswa dalam menyiapkan tugas mereka berkaitan penggunaan pantun dengan menggunakan teknik tersendiri untuk meramal penggunaan pantun yang betul dalam ayat. Secara tidak langsung, masa yang diperlukan untuk menelaah atau mencari pantun yang diinginkan dapat diijimatkan dengan adanya sistem ini. Hal ini juga bertindak sebagai inovasi kepada golongan mahasiswa untuk lebih berinteraksi dengan komponen digital seperti sistem ini, dan tidak terhad kepada rujukan ilmiah berasaskan kertas semata-mata sahaja.

## **RUJUKAN**

Gonzalez, Z. G., Kelly, M., Murphy Jr, T. E. & Nisenson, M. 2013. Search Engine Indexing, Google Patents.

Theobald, A. & Weikum, G. 2002. The Index-Based Xxl Search Engine for Querying Xml Data with Relevance Ranking. International Conference on Extending Database Technology, hlm. 477-495.

Croft, W. B., Metzler, D. & Strohman, T. 2010. Search Engines: Information Retrieval in Practice. 283. Addison-Wesley Reading.

Nik Safiah Karim. 2006. *Tatabahasa Dewan*. Kuala Lumpur. Dewan Bahasa dan Pustaka.

Pusat Rujukan Persuratan Melayu, Dewan Bahasa dan Pustaka

<http://prpmv1.dbp.gov.my/Search.aspx?k=pantun>

Laman web rasmi Dewan Bahasa dan Pustaka

<http://laman.dbp.gov.my/lamandbp/main.php?Content=vertsections&SubVertSectionID=826&VertSectionID=25&CurLocation=208&IID=&Page=1>

Rouse, M. 2015. parser. <http://searchmicroservices.techtarget.com/definition/parser> [April 2005]

Raka. 2015. Koleksi Pantun Terbaru.

<https://play.google.com/store/apps/details?id=com.raka.koleksipantunterbaru>

[12March 2015]

Nurmalabs7. 2016. Pantun dan Puisi.

<https://play.google.com/store/apps/details?id=com.nurmalabs7.pantundanpuisien> [12

April 2013]s

Aisydev. 2014. Pantun Lucu Jenaka.

<https://play.google.com/store/apps/details?id=com.Aisydev.pantunlucujenaka&hl=in>

[19 Februari 2017]

Prau Media. 2017. Pantun Pendidikan.

<https://play.google.com/store/apps/details?id=com.praumedia.pantunpendidikan>

[19 Februari 2017]

Copyright@FTSM