

<http://www.ftsm.ukm.my/apjitm>

Asia-Pacific Journal of Information Technology and Multimedia

Jurnal Teknologi Maklumat dan Multimedia Asia-Pasifik

Vol. 2 No. 1, June 2013 : 27 - 37

e-ISSN: 2289-2192

KEBARANGKALIAN SECARA AUTOMATIK DALAM ALKHWARIZMI BAYES UNTUK APLIKASI AGEN BUALAN

NAZLIA OMAR
MASNIZAH MOHD
YUSMAN JAMAT

ABSTRAK

Agen bualan merupakan salah satu aplikasi yang boleh diguna untuk mendapat maklumat secara dialog mengguna bahasa tabii. Kajian ini mencadangkan alkhwarizmi Bayes yang mengguna nilai kebarangkalian yang dijana secara automatik untuk mentafsir kehendak pengguna kerana dialog ada kalanya membawa maksud yang samar atau mengandungi maklumat yang tidak lengkap. Teknik ini diaplikasi ke dalam agen bualan untuk melihat keberkesanan kaedah yang dicadang. Agen bualan yang dihasil mengguna persekitaran Bahasa Melayu sebagai bahasa interaksi. Selain itu, agen yang dihasil mampu menganalisis keperluan individu tanpa memerlukan input yang lengkap daripada pengguna berdasarkan analisis dan pengemaskinian nilai kebarangkalian yang digunakan dalam pengiraan alkhwarizmi Bayes.

Kata kunci: Alkhwarizmi Bayes, Padanan Corak, Agen Bualan

ABSTRACT

Conversational agent is an application that can be used to obtain information using natural language processing. This study proposes Bayes algorithm using probability value generated automatically to interpret user requirement. User dialogue involves ambiguous meaning and sometimes occur an incomplete information. Bayes algorithm is applied to conversational agent to investigate the effectiveness of the proposed method. Conversational agents produced are using Malay environment as an interaction language. In addition, the conversational agents are able to analyze user need without having a complete input. It is achievable based on the analysis and updated probability value used in the calculation of Bayes algorithm.

Keywords: Bayes Algorithm, Pattern Matching, Conversational Agent

PENGENALAN

Agen bualan merupakan satu sistem yang membekal maklumat kepada pengguna dalam bentuk dialog. Capaian maklumat dan memetik maklumat adalah mengenai mencari jawapan berdasarkan soalan tertentu (Bayan et al. 2004). Jacques (2005) pula menyatakan, perbualan adalah rangka kerja umum untuk pengurusan struktur proses pembelajaran yang baik. Sistem ini memahami kehendak pengguna dan memberi perkhidmatan yang diperlu oleh pengguna dengan mengguna dialog yang diterima. Capaian maklumat dalam pangkalan data oleh sistem memerlukan input yang lengkap daripada pengguna dan ini menimbulkan masalah apabila berlaku kesamaran pada input. Untuk mengatasi kelemahan ini, kajian ini mencadangkan penggunaan alkhwarizmi Bayesian dengan mengguna nilai kebarangkalian yang dijana secara automatik.

Perbualan dalam talian semakin meningkat penggunaannya dalam kalangan pengguna rangkaian internet dan komputer. Perbualan dalam talian merujuk kepada komunikasi melalui talian internet, memfokus kepada perbualan *one-on-one* dengan mengguna perkakasan seperti *Internet Relay Chat* (IRC) (Aida 2004). Secara amnya, perbualan dalam talian merupakan satu

aktiviti perbualan antara dua pihak atau pengguna dengan mengguna komputer dan rangkaian internet iaitu pengguna berada pada kedudukan yang jauh. Perbualan dilakukan dengan menaip perkataan atau teks mengguna komputer untuk dihantar melalui rangkaian internet dan dibaca oleh pihak yang berada di tempat lain. Aktiviti perbualan dalam talian ini menjadi satu aktiviti masa nyata iaitu perkataan yang ditaip dan dihantar boleh terus dibaca dan dibalas oleh pihak yang lain dalam masa yang singkat. Kebanyakan isi perbualan memfokus topik tertentu yang menarik minat yang sama antara kedua pihak yang melakukan aktiviti perbualan tersebut. Antara aplikasi perisian dan laman web yang membolehkan aktiviti perbualan dalam talian ialah mIRC, Yahoo Messenger, MSN Messenger, Facebook, Twitter dan Friendster.

Perisian perbualan dalam talian menjadi perisian penghubung perbualan antara manusia dan manusia kini berkembang dengan inovasi agen bualan yang memfokus kepada interaksi manusia-mesin. Perbualan dan corak pertuturan manusia dilaksanakan ke bentuk alkhwarizmi dalam sistem agen bualan seperti *chatbot* Eliza. Agen bualan merupakan antara muka sistem yang mewakili seseorang pengguna. Maklumat berguna disampaikan kepada pengguna sistem melalui beberapa kaedah seperti suara, pergerakan dan teks. Agen bualan mengeksploitasi teknologi bahasa semula jadi untuk berhubung dengan pengguna berasaskan teks (James et al. 2004). Agen bualan merupakan satu aplikasi yang memberi perkhidmatan yang membekal maklumat pengguna mengikut keperluan dan dalam bentuk perbualan antara pengguna dan komputer. Perkhidmatan ini sering diguna dalam bidang perniagaan, kerajaan, pendidikan, kesihatan dan hiburan. Malah agen bualan menjadi satu sistem yang mampu melakukan aktiviti perbualan antara sistem dan pengguna dengan tujuan untuk membekal maklumat kepada pengguna dan menarik pengguna untuk terus mengguna sistem.

AGEN BUALAN

Agen bualan atau dahulunya dikenali sebagai *chatbot* merupakan sistem yang berinteraksi dengan pengguna melalui pendekatan bahasa tabii. Pelbagai aplikasi yang mengguna agen bualan seperti perkhidmatan pelanggan dan pusat panggilan semakin berkembang. Penggunaan komputer peribadi secara meluas meningkat lagi keperluan hubungan interaksi yang baik antara komputer dan pengguna. Dialog merupakan satu inisiatif untuk menunjuk minat, permintaan atau kueri secara terus secara kebiasaannya melalui pertuturan, menaip dan menunjuk (Zadrozny et al. 2000).

Agen bualan mula diperkenal pada tahun 1960-an. Agen bualan Eliza merupakan antara agen bualan yang terawal pada tahun 1966. Ia dihasil daripada kajian ke atas pemprosesan bahasa tabii dengan menggunakan teknik padanan mudah (Lee & Cho 2001). Teknologi Eliza menjurus kepada pemprosesan bahasa tabii. Ia mengguna teknik padanan corak dalam mencapai maklumat. Sebagai contoh, jika pengguna memasukkan input yang mengandungi perkataan “nama”, Eliza menjawab “Nama saya Eliza” yang mana jawapan mempunyai perkataan “nama” yang sepadan dengan input.

ALICE dihasil pada tahun 1995 oleh Richard Wallace daripada Universiti Lehigh dan dikenali sebagai AliceBot. ALICE dibangun dengan mengguna AIML (*Artificial Intelligence Markup Language*) yang berasaskan XML. Reka bentuk ALICE sama dengan agen bualan yang lain sejak ELIZA diperkenal. Kelebihan ALICE ialah enjin yang mempunyai carta Graf Induk (*Graphmaster*) iaitu mencarta semua kategori untuk memudah carian corak dan padanan. Alkhwarizmi yang diguna masih sama seperti ELIZA yang mengguna padanan kata kunci dan menjawab dengan set peraturan yang ditetapkan. Bagaimanapun, “otak” atau data tidak boleh diubah suai kerana ALICE tidak mempunyai kebolehan untuk belajar. ALICE dilengkapi dengan log fail untuk merekod perbualan, mengingat pengguna dan untuk menentu topik semasa sesi perbualan seterusnya. Perbezaan besar antara ALICE dan ELIZA terletak pada pengurusan fail. Fail ALICE berstruktur kerana AIML direka untuk bagi menghasil agen

bualan. Manakala untuk padanan corak, ALICE mengguna beberapa teknik seperti *wildcard*, penormalan input dan pemetaan graf.

Sejak beberapa dekad, teknologi agen semakin berkembang dari semasa ke semasa. Kajian mengenai agen semakin banyak dihasil sebagai alat untuk memudah pembelajaran pengguna dengan cara semula jadi dan mampu memberi motivasi. Agen memperoleh maklumat baharu melalui beberapa kaedah seperti memproses bahasa tabii, memantau tindakan pengguna atau belajar dari pembetulan oleh pengguna atau jurulatih selepas kesilapan dilakukan oleh agen (Aida 2004).

Ai (2010) menjelaskan HAL sebagai kanak-kanak maya iaitu program perisian yang mempunyai kebolehan mendapat pengetahuan seperti kanak-kanak dengan cara latihan dan amalan, cuba dan gagal. HAL dibangun berdasar prinsip Alan Turing sejak setengah abad yang lalu, menghasil kecerdasan buatan yang sebenar iaitu “mesin kanak-kanak”, alkhwarizmi belajar yang mempunyai kebolehan belajar melalui pengalaman, boleh diajar bertutur sama seperti perkembangan yang dilihat pada bayi yang membesar.

Satu lagi agen yang dikategori sebagai agen pintar ialah agen yang mengguna pendekatan *Wizard-of-Oz*. Teknik ini membolehkan agen memperoleh kebolehan perbualan secara dasarnya dengan mengumpul contoh dialog dan model dialog belajar (Okatomo, M. 2002). Model dialog yang lengkap tidak dapat diperolehi dari awal kerana kesukaran mengumpul dialog berkaitan dengan topik tertentu. *Wizard-of-Oz* menggunakan model dialog alkhwarizmi peningkatan pembelajaran dalam mengumpul dialog yang boleh diguna untuk perbualan.

Sejarah agen bualan mula diperkenal sejak tahun 1960-an dengan ELIZA. Semenjak itu, teknologi agen bualan semakin berkembang dan penggunaannya semakin meluas. Dari penggunaan padanan corak, agen bualan berevolusi kepada kecerdasan buatan yang bertindak dan melakukan aktiviti tanpa perlu arahan secara terus daripada pengguna. Pelbagai pembaharuan telah diperkenal untuk meningkatkan teknologi agen bualan. Antaranya ialah penggunaan audio, video, pergerakan badan dan tindak balas spontan agen terhadap input yang diterima. Selain daripada itu pelbagai teknik baharu diaplikasi untuk menjadikan perbualan antara agen bualan dan pengguna menjadi realistik dan semula jadi. Antara teknik yang diguna dalam agen bualan adalah alkhwarizmi Bayes, rangkaian Bayes, *Wizard-of-Oz* dan HAL. Bagaimanapun, kebanyakan agen bualan yang pernah dihasil mengguna bahasa asing selain daripada Bahasa Melayu sebagai medium interaksi.

TEKNIK PADANAN

Teknik padanan merupakan antara teknik tradisional yang memilih kata kunci daripada input pengguna dan membanding kata kunci yang terdapat dalam pangkalan data (Kim et al. 2006). Nilai yang tinggi diberi apabila terdapat padanan kata kunci. Contoh padanan kata kunci boleh dilihat dalam Jadual 1 dengan mengguna rumus kejituan, dapatan dan *f-measure* yang diadaptasi dari hasil kerja (Kim, et al. 2006).

$$\text{Kejituan} = \frac{\text{Jumlah kata kunci berpadanan Corak}}{\text{Jumlah bilangan kata kunci kueri}}$$

$$\text{Dapatan} = \frac{\text{Jumlah kata kunci berpadanan Corak}}{\text{Jumlah kata kunci Corak}}$$

$$F - \text{measure} = \frac{(\alpha + 1) \times \text{ketepatan} \times \text{recall}}{\alpha \times \text{ketepatan} + \text{recall}}, \alpha = 1$$

JADUAL 1. Contoh Padanan Kata Kunci

<i>Kueri</i>	<i>Kata kunci</i> (<i>v w x y</i>)	<i>Padanan Tradisional</i> (<i>kejituan</i>)	<i>Dapatan</i>	<i>F-measure (a:1)</i>
Corak A	v w x	3/4 = 0.75 (X)	3/3 = 1.0 (✓)	0.86 (X)
Corak B	v w x y z	4/4 = 1.0 (✓)	4/5 = 0.8 (X)	0.89 (X)
Corak C	v w x y	4/4 = 1.0 (✓)	4/4 = 1.0 (✓)	1.0 (✓)

Semakin banyak maklumat dalam pangkalan data, terdapat tindanan atau corak yang hampir sama seperti yang dilihat dalam Jadual 1. Jumlah maklumat perlu diambil kira. *F-measure* diguna untuk mengatasi masalah pengkelasan teks. Berdasarkan Jadual 1, corak C adalah yang paling hampir atau paling sepadan dengan kata kunci ‘v w x y’ berdasarkan nilai kejituan, dapatan dan *f-measure*.

Padanan corak secara tradisional merupakan kaedah yang sesuai diguna apabila input yang dimasukkan oleh pengguna sepadan dengan data yang disimpan dalam pangkalan data. Walaupun begitu padanan corak secara tradisional mempunyai satu kelemahan yang nyata iaitu padanan antara input dan data tidak dapat dibuat apabila input yang diterima daripada pengguna tidak lengkap atau melibatkan sebahagian daripada data yang disimpan dalam pangkalan data.

Bagi mengatasi kelemahan padanan corak tradisional, satu alkhwarizmi telah diguna. Kajian ini mencadangkan penggunaan kebarangkalian yang dijana secara automatik untuk pengiraan alkhwarizmi Bayes dalam mentafsir kehendak pengguna walaupun input yang diterima daripada pengguna tidak lengkap. Rumus alkhwarizmi Bayes seperti yang dinyatakan oleh Po (2006) adalah:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|not A)P(not A)}$$

Berdasarkan data rekod, kebarangkalian berdasarkan data tersebut boleh dihasil. Jadual 2 menunjukkan contoh objek dan atribut dalam pangkalan data.

JADUAL 2. Objek dan Atribut Data

<i>Objek</i>	<i>Atribut</i>
Kakitangan Akademik	nama (Omar Ali, Abu Bakar), jawatan akademik, jawatan pentadbiran, emel.
Kakitangan Sokongan	nama (Abu Bakar Omar), jawatan, emel.

Berdasarkan objek dan atribut Jadual 2, kebarangkalian untuk setiap nama (lajur pertama) dan atribut dapat dihasil seperti dalam Jadual 3.

JADUAL 3. Kebarangkalian daripada nama (lajur pertama setiap jadual)

<i>Nama</i>	<i>Bil.</i>	<i>P(T)</i>
Abu Bakar Omar	1	$\frac{1}{3} = 0.33$
Abu Bakar	1	$\frac{1}{3} = 0.33$
Omar Ali	1	$\frac{1}{3} = 0.33$

<i>Abu</i>	<i>Bakar</i>	<i>Omar</i>	<i>P(T)</i>
0	0	0	0.0
0	0	1	0.33
0	1	0	0.33
0	1	1	0.67
1	0	0	0.33

1	0	1	0.67
1	1	0	0.67
1	1	1	1.0

<i>Abu</i>	<i>Bakar</i>	<i>P(T)</i>
0	0	0.0
0	1	0.5
1	0	0.5
1	1	1.0

<i>Omar</i>	<i>Ali</i>	<i>P(T)</i>
0	0	0.0
0	1	0.5
1	0	0.5
1	1	1.0

Jadual 3 menunjukkan pengiraan kebarangkalian yang dilakukan ke atas setiap objek yang terdapat dalam jadual. Objek dalam kes ini menggunakan nama (lajur-pertama) yang terdapat dalam jadual yang diguna untuk pengiraan alkhwarizmi Bayes pada peringkat pertama.

JADUAL 4. Kebarangkalian daripada atribut (objek Kakitangan Akademik)

<i>Atribut</i>	<i>Bil</i>	<i>P(T)</i>
nama	1	0.25
jawatan akademik	1	0.25
jawatan pentadbiran	1	0.25
emel	1	0.25

<i>nama</i>	<i>P(T)</i>
0	0.0
1	1.0

<i>emel</i>	<i>P(T)</i>
0	0.0
1	1.0

<i>jawatan</i>	<i>akademik</i>	<i>P(T)</i>
0	0	0.0
0	1	0.5
1	0	0.5
1	1	1.0
0	0	0.0
0	1	0.5
1	0	0.5
1	1	1.0

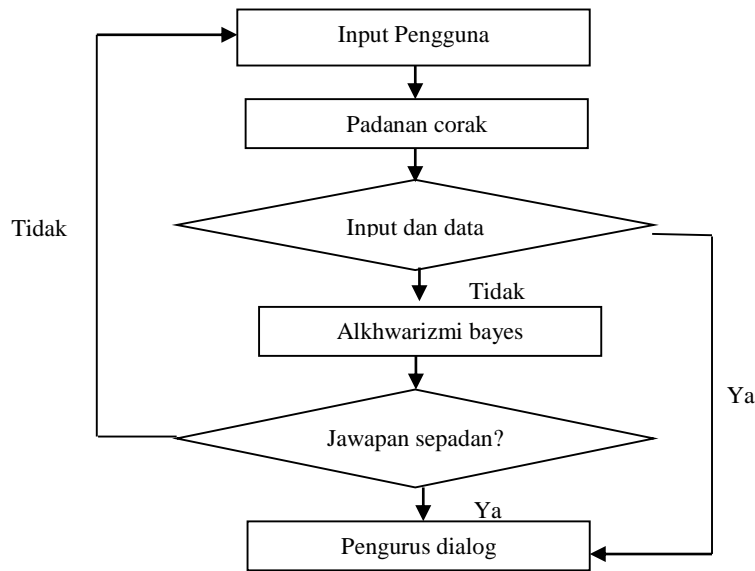
Jadual 4 dihasil berdasarkan atribut yang terdapat dalam jadual Kakitangan Akademik. Atribut yang terdapat dalam jadual kakitangan akademik ialah nama, jawatan akademik, jawatan pentadbiran dan emel. Kebarangkalian atribut untuk objek Kakitangan Sokongan juga dihasil sama seperti Jadual 4. Atribut yang terdapat dalam jadual Kakitangan Sokongan adalah nama, jawatan dan emel. Setiap nilai kebarangkalian yang diperolehi daripada atribut diguna sebagai nilai awalan untuk pengiraan alkhwarizmi Bayes pada peringkat kedua.

SENIBINA

Agensi bualan ini mencadangkan alkhwarizmi Bayes dengan menggunakan nilai kebarangkalian yang dihasil secara automatik. Ini kerana nilai kebarangkalian alkhwarizmi Bayes selalunya dihasil daripada kajian atau melalui pendapat pakar. Walaupun begitu, hasil kajian atau pendapat pakar mungkin kurang tepat mengikut keperluan pengguna yang berbeza dalam

mencapai data. Kajian ini mencadangkan nilai kebarangkalian automatik diguna sebagai nilai permulaan dan nilai tersebut dikemas kini mengikut keperluan pengguna.

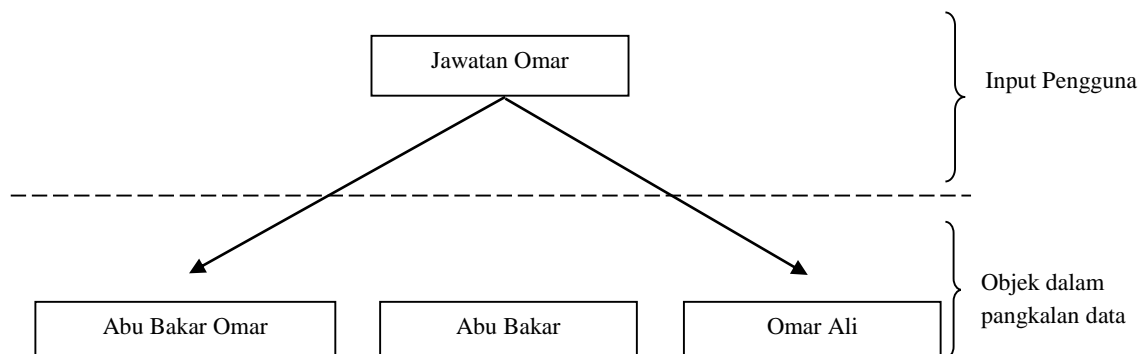
Rajah 1 merupakan reka bentuk agen bualan yang mengguna teknik padanan corak dan alkhwarizmi Bayes untuk mentafsir kehendak pengguna. Jika teknik padanan corak antara input dan data gagal dilakukan, agen mengguna alkhwarizmi Bayes untuk mentafsir kehendak pengguna. Pengurus dialog diguna untuk meminta input tambahan jika carian gagal dilakukan dan seterusnya memapar maklumat kepada pengguna.



RAJAH 1. Aliran Data Agen Bualan

Proses carian maklumat akan melalui dua proses iaitu padanan corak tradisional dan seterusnya mentafsir mengguna alkhwarizmi Bayes jika padanan corak gagal dilakukan. Alkharizmi Bayes melakukan pengiraan sebanyak dua peringkat. Jika input pengguna adalah “jawatan Omar” maka:

Peringkat pertama merupakan pengiraan nama (lajur pertama) untuk setiap objek yang ada dalam jadual mengguna nilai kebarangkalian dalam Jadual 3.



RAJAH 2. Gambaran hubungan antara input dan objek dalam pangkalan data

Abu Bakar Omar = D1

$P(D1 = t \mid \text{Abu} = 0, \text{Bakar} = 0, \text{Omar} = 1)$

Jika $T = (0 \ 0 \ 1)$

$$\begin{aligned}
 P(D1 = t \mid T = t) &= \frac{P(T=t \mid D1=t) P(D1=t)}{P(T = t \mid D1 = t)P(D1=t) + P(T = t \mid D1 = f)P(D1=f)} \\
 &= \frac{0.33 \times 0.33}{(0.33 \times 0.33) + (0.67 \times 0.67)} = \frac{0.11}{0.56} = 0.2
 \end{aligned}$$

Abu Bakar = D2

$P(D2 = t \mid \text{Abu} = 0, \text{Bakar} = 0)$

Jika $T = (0 \ 0)$

$$P(D2 = t \mid T = t) = \frac{P(T=t \mid D2=t) P(D2=t)}{P(T = t \mid D2 = t)P(D2=t) + P(T = t \mid D2 = f)P(D2=f)}$$

$$= \frac{0 \times 0.33}{(0 \times 0.33) + (1 \times 0.67)} = \frac{0}{0.67} = 0.0$$

Omar Ali = D3

$P(D3 = t \mid \text{Omar} = 0, \text{Ali} = 1)$

Jika $T = (0 \ 1)$

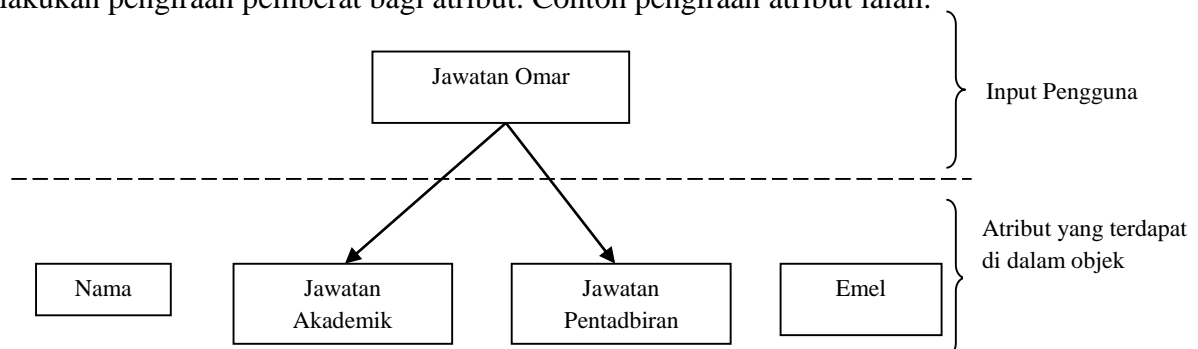
$$P(D3 = t \mid T = t) = \frac{P(T=t \mid D3=t) P(D3=t)}{P(T = t \mid D3 = t)P(D3=t) + P(T = t \mid D3 = f)P(D3=f)}$$

$$= \frac{0.5 \times 0.33}{(0.5 \times 0.33) + (0.5 \times 0.67)}$$

$$= \frac{0.166}{0.5} = \underline{0.33}$$

Rajah 2 menunjukkan “Abu Bakar” tidak ada hubungan dengan maklumat dalam pangkalan data dan seterusnya memberi nilai pemberat 0.0. Berdasarkan hasil pengiraan peringkat pertama, “Omar Ali” dipilih oleh agen bualan kerana mempunyai nilai pemberat yang tertinggi.

Peringkat kedua merupakan kesinambungan daripada peringkat 1, “Omar Ali” merupakan maklumat yang terdapat dalam objek Kakitangan Akademik. Jadual 4 diguna untuk melakukan pengiraan pemberat bagi atribut. Contoh pengiraan atribut ialah:



RAJAH 3. Gambaran hubungan antara input dan atribut dalam objek

nama = k1

$P(k1 = t \mid \text{nama} = 0)$

Jika $T = (0)$

$$P(k1 = t \mid T = t) = \frac{P(T=t \mid k1=t) P(k1=t)}{P(T = t \mid k1 = t)P(k1=t) + P(T = t \mid k1 = f)P(k1=f)}$$

$$= \frac{0 \times 0.25}{(0 \times 0.25) + (1 \times 0.75)} = \frac{0}{0.75} = 0.0$$

jawatan akademik = k2

$P(k2 = t \mid \text{jawatan} = 1, \text{akademik} = 0)$

Jika $T = (1 \ 0)$

$$P(k2 = t \mid T = t) = \frac{P(T=t \mid k2=t) P(k2=t)}{P(T = t \mid k2 = t)P(k2=t) + P(T = t \mid k2 = f)P(k2=f)}$$

$$= \frac{0.5 \times 0.25}{(0.5 \times 0.25) + (0.5 \times 0.75)} = \frac{0.125}{0.5} = \underline{0.25}$$

$$\begin{array}{ll}
\text{jawatan pentadbiran} = k3 & \text{emel} = k4 \\
P(k3 = t \mid \text{jawatan} = 1, \text{pentadbiran} = 0) & P(k4 = t \mid \text{emel} = 0) \\
\text{Jika } T = (1 \ 0) & \text{Jika } T = (0) \\
P(k3 = t \mid T = t) = \underline{0.25} & P(k4 = t \mid T = t) = 0.0
\end{array}$$

Nilai pemberat yang diberi kepada “nama” dan “emel” ialah 0.0 kerana tiada hubungan secara langsung dengan input pengguna. Berdasarkan hasil pengiraan peringkat kedua, “jawatan akademik” dan “jawatan pentadbiran” dipilih oleh agen bualan kerana mempunyai nilai pemberat yang paling tinggi.

KEPUTUSAN UJIAN

Beberapa peraturan ditetapkan untuk alkhwarizmi padanan corak (PCT) tradisional dan gabungan padanan corak tradisional bersama alkhwarizmi Bayes (PCT+BA) yang diuji. Ini memastikan ujian yang dilakukan menghasil keputusan yang sah. Peraturan yang ditetapkan untuk melaksana eksperimen ini adalah:

1. input terdiri daripada istilah yang terdapat dalam nama
2. hanya satu istilah sahaja yang dimasukkan setiap kali input
3. setiap istilah hanya dimasukkan sekali sahaja untuk melihat tindak balas sistem

Ujian ini dilakukan dengan memasukkan sebanyak 302 istilah iaitu setiap istilah terdapat dalam nama yang disimpan dalam pangkalan data. Satu jadual dihasil iaitu Jadual 5, kesimpulan ujian output PCT dan PCT+BA. Walaupun begitu bilangan output relevan yang mempunyai nilai 1 tidak dimasukkan dalam Jadual 5 kerana tidak memapar keberkesanan mana-mana teknik dalam memilih dan mentafsir data mengikut keperluan input.

JADUAL 5. Kesimpulan Ujian Output PCT dan PCT+BA

Bil	Input	Bilangan Output		
		<i>PCT</i>	<i>PCT+BA</i>	<i>Relevan</i>
1	ab	0	2	2
2	abd	0	1	3
3	abdul	0	4	10
4	abdullah	0	4	6
5	abu	0	1	3
6	ahmad	0	2	7
7	ali	0	4	4
8	arif	0	1	2
9	aziz	0	1	3
10	azlina	0	1	2
11	bakar	0	2	2
12	hakim	0	1	2
13	ishak	0	1	2
14	lokman	0	1	2
15	md	0	5	6
16	mohamad	0	3	3
17	mohamed	0	2	3

18	mohammad	0	2	3
19	mohd	0	1	25
20	mokhtar	0	2	2
21	muhamad	0	1	2
22	noh	0	1	2
23	noor	0	1	4
24	nor	0	2	4
25	noraidah	0	1	2
26	nur	0	2	3
27	omar	0	1	2
28	osman	0	2	2
29	othman	0	1	2
30	rahman	0	3	3
31	rashid	0	1	2
32	razak	0	1	2
33	razali	0	2	2
34	siti	0	1	5
35	sulaiman	0	1	2
36	yahya	0	2	2
37	yusof	0	3	3
38	zainal	0	2	3
39	zakaria	0	2	2
40	zin	0	1	2
41	zulkarnain	0	1	2

Tidak ada output dipapar oleh agen yang mengguna alkhwarizmi padanan corak tradisional sahaja ini kerana bilangan input tidak mencukupi untuk PCT melakukan padanan antara input dan data yang disimpan. Satu lagi analisis dilakukan untuk melihat keberkesanan alkhwarizmi PCT+BA. Analisis tersebut boleh dilihat pada Jadual 6.

JADUAL 6. Analisa Keberkesanan Alkhwarizmi PCT+BA

Bilangan		
<i>output = 1</i>	<i>1 < output < relevan</i>	<i>output = relevan</i>
21	9	11

Jadual 6 menunjukkan alkhwarizmi yang dicadangkan dapat menyenarai pendek output berdasarkan input yang dimasukkan. Ia menunjukkan 51% daripada output memapar satu output sahaja. 22% daripada output menunjukkan lebih dari satu output dipapar, tetapi masih kurang daripada output yang relevan. Selebihnya sebanyak 27% memapar output yang sama dengan output yang relevan.

PERBINCANGAN

Nilai kebarangkalian yang dijana secara automatik mempunyai kebolehan untuk mentafsir kehendak pengguna. Walaupun begitu, nilai kebarangkalian tersebut menjadi nilai awalan yang diguna untuk mencari maklumat pada peringkat awalan. Ini bermakna, mungkin terdapat lebih

daripada satu maklumat yang sepadan atau mempunyai pemberat yang tinggi. Ini bererti lebih daripada satu kata kunci boleh diguna untuk mendapat output yang tepat.

Kaedah lain yang boleh diguna untuk meningkat kejituan capaian maklumat ialah dengan mengemas kini kebarangkalian apabila pengguna membuat carian. Sebagai contoh pengiraan alkhwarizmi Bayes pada peringkat kedua, jika pengguna memilih “jawatan akademik”, Jadual 4 dikemas kini dan menjadi Jadual 7.

JADUAL 7. Kebarangkalian daripada atribut (objek Kakitangan Akademik) selepas kemas kini

Atribut	Bil	P(T)
nama	1	0.2
jawatan akademik	2	0.4
jawatan pentadbiran	1	0.2
emel	1	0.2

nama	P(T)
0	0.0
1	1.0

emel	P(T)
0	0.0
1	1.0

jawatan	akademik	P(T)
0	0	0.0
0	1	0.5
1	0	0.5
1	1	1.0

jawatan	pentadbiran	P(T)
0	0	0.0
0	1	0.5
1	0	0.5
1	1	1.0

Berdasarkan Jadual 7, kebarangkalian jawatan akademik adalah tinggi iaitu 0.4 berbanding dengan yang lain iaitu 0.2. Ini adalah kerana selepas pengguna memilih “jawatan akademik”, agen menganggap bahawa “jawatan akademik” ialah maklumat yang diperlu pengguna jika pengguna memasukkan input yang sama iaitu “jawatan”. Berdasar nilai kebarangkalian selepas pengemaskinian di Jadual 7, jika agen masih menerima “jawatan omar” sebagai input daripada pengguna, “jawatan akademik” mendapat nilai pemberat yang paling tinggi selepas pengiraan alkhwarizmi Bayes peringkat kedua diaplikasi iaitu 0.4.

Penggunaan alkhwarizmi Bayes dengan nilai kebarangkalian secara automatik dan pengemaskinian nilai kebarangkalian berdasar keperluan pengguna menjadikan agen bualan perlu bijak mentafsir kehendak pengguna. Ini kerana pengemaskinian kebarangkalian dilakukan berdasar carian maklumat yang dilakukan oleh pengguna tersebut. Ini menjadikan agen bualan dapat mentafsir input yang menjurus kepada keperluan individu, berbanding nilai kebarangkalian berdasarkan kajian atau pendapat pakar dengan hasil nilai kebarangkalian diperolehi daripada kelompok tertentu sahaja.

KESIMPULAN

Agan bualan ialah satu antara muka kompleks yang menghubungkan pengguna dan sistem dan berkomunikasi dalam bentuk dialog. Penggunaan Bahasa Melayu sebagai medium interaksi antara agen bualan dan pengguna menjadi satu keperluan memandangkan teknologi ini semakin

berkembang dan penggunaannya semakin meluas. Kajian dilakukan untuk menghasil sebuah agen yang mengaplikasi Bahasa Melayu dengan mengguna teknik yang dipilih dan diubah suai untuk mentafsir kehendak pengguna. Alkharizmi yang dipilih untuk mentafsir kehendak pengguna adalah alkharizmi Bayes.

Alkharizmi Bayes dengan kebarangkalian yang dijana secara automatik boleh diaplikasi dalam agen bualan sebagai alkharizmi yang mampu mentafsir kehendak pengguna. Nilai kebarangkalian yang dijana menjadi nilai permulaan agen bualan. Pengemaskinian nilai kebarangkalian berdasar carian maklumat oleh pengguna menjadikan tafsiran kehendak pengguna menjadi baik dan tepat. Kaedah penggunaan alkharizmi Bayes dalam agen bualan menjadi baik kerana nilai kebarangkalian secara automatik dan pengemaskinian kebarangkalian yang dilakukan adalah berdasar keperluan individu yang mengguna agen bualan.

RUJUKAN

- Ai. 2010. Who is HAL. http://www.a-i.com/show_tree.asp?id=97 &level=2&root=115 [22 April 2010].
- Aida Mustapha. 2004. *Learnable Conversational Agent: A Framework for Home-Based Education*. Tesis Sarjana, Universiti Kebangsaan Malaysia.
- Bayan, A.S. and Eric, A. 2004. Accessing an Information System by Chatting. *Natural Language Processing and Information Systems*, 3136:407-412.
- Jacques, L. 2005. A structured chat framework for distributed educational settings. *Proceedings of International Society of the Learning Sciences*. Taiwan: IEEE Computer Society, 403-407.
- James, L., Karl, B. and Bradfort, M. 2004. *Practical Handbook of Internet Computing*. London: Chap & Hall.
- Kim, K-Y, Hong, J-H, Cho, S-B. 2006. A Semantic Bayesian Network Approach to Retrieving Information with Intelligent Conversational Agents. *Information Processing and Management*, 43(1):225-236.
- Lee, S-I and Cho, S-B. 2001. An Intelligent With Structured Pattern Matching For A Virtual Representative. *Intelligent Agent Technology, World Scientific*, 305 – 309.
- Okamoto, M. 2002. Incremental PDFA Learning for Conversational Agents. *IEEE International Workshop on Knowledge Media Networking (KMN-02), Kyoto, Japan*, 161-166.
- Po, I. 2006. An Approach of The Naive Bayes Classifier for the Document Classification. *General Mathematics*, 14(4):135-138.
- Richard, S.W. 2003. *The Elements of AIML Style*. [s.l.]:ALICE A.I. Foundation, Inc.
- Zadrozny, W., Budzikowska, M., Chai, J., Kambhatla, N., Levesque, S., and Nicolov, N. 2000. Natural Language Dialogue for Personalized Interaction. *Commun. ACM* 43(8):116 – 120.

Nazlia Omar,
Masnizah Mohd,
Yusman Jamat
Fakulti Teknologi & Sains Maklumat
Universiti Kebangsaan Malaysia
43600 Bangi, Selangor, Malaysia
no@ftsm.ukm.my, mas@ftsm.ukm.my, yusman_jamat@yahoo.com