Advanced Journal of Technical and Vocational Education 1 (1): 01-05, 2017
ISSN:

DOI:

Sistem Soal Jawab Menggunakan Model Ruang Vektor Bagi Ibadah Saie

Azura bt Ahmad *[1], Nurul Ihsaniah bt Omar [2], Nur Fairuz Afni bt Ahmad Faizal [3]

[1] Azura bt Ahmad, Politeknik Balik Pulau, Jabatan Teknologi Maklumat dan Komunikasi, Balik Pulau, Penang, Malaysia

[2] Nurul Ihsaniah bt Omar, Politeknik Balik Pulau, Jabatan Teknologi Maklumat dan Komunikasi, Balik Pulau, Penang, Malaysia

ARTICLE INFO

Article history:
Received
Accepted
Available online

Keywords:
Capaian maklumat, QA(Soal-Jawab), VSM (Model Ruang Vektor), Persamaan padanan, Kejituan, Dapatan

ABSTRACT

Kajian ini untuk menilai kerelevanan kaedah Model Ruang Vektor (VSM) untuk diaplikasikan ke dalam Sistem Soal-Jawab Ibadah Saie. Reka bentuk model kajian, menggunakan teknik beg perkataan (BOW), soalan faktoid jenis WH dan VSM iaitu pemberat perkataan dan persamaan kosinus. Kejituan dan Dapatan digunakan bagi mendapatkan penilaian ketepatan pemangkatan capaian jawapan. Hasil pemangkatan dokumen daripada Ujian Tertutup (set latihan) dan Ujian Terbuka (set ujian) menunjukkan pendekatan VSM boleh diaplikasikan ke dalam sistem tersebut**.**

PENGENALAN

Dalam teknologi Capaian Maklumat (IR), *Frequently Ask Question* (FAQ) merupakan satu sistem yang boleh digunakan untuk mencapai dokumen menggunakan bahasa tabii sebagai kueri dalam proses capaian maklumat relevan. Sistem IR memproses dan menyimpan maklumat yang tidak berstruktur dalam jumlah kuantiti yang besar, di mana ia tidak mempunyai satu format tertentu (biasanya teks) di dalam turutan yang efisyen, maka ia mampu memberi respon pantas yang relevan dengan kueri yang diberi [3]. Teknologi FAQ dalam sistem QA dapat menganalisis soalan-soalan daripada pengguna, mencari kueri yang mempunyai persamaan tertinggi dengan soalan-soalan yang terdapat di dalam FAQ dan jawapan yang terkandung di dalam pangkalan data berkeupayaan untuk memberi jawapan yang memenuhi kehendak pengguna [4].

Kajian ini menumpukan kepada capaian maklumat menggunakan kaedah sistem QA untuk subjek Saie yang terdapat dalam rukun mengerjakan ibadah Haji dan Umrah bagi membantu memudahkan pengguna terutamanya Muslim mendapatkan maklumat yang lebih fokus. Dokumen korpus Saie dan set soalan yang diproses adalah dalam Bahasa Inggeris. Secara khususnya kajian ini menumpukan kepada, menganalisis VSM di dalam pembangunan sistem QA subjek Saie di mana sistem akan mengekstrak perkataan berdasarkan kueri pengguna dan padankan dengan soalan yang terdapat di dalam domain, seterusnya sistem memberi capaian pangkatan jawapan yang relevan dengan kueri pengguna.

MODEL RUANG VEKTOR

VSM merupakan asas kepada operasi capaian maklumat yang banyak termasuklah memperolehi dokumen-dokumen pada satu pertanyaan, pengelasan dokumen dan pengklusteran dokumen [1]. Sistem QA ibadah Saie merupakan satu aplikasi sistem capaian maklumat secara automatik yang mengimplementasikan pendekatan VSM untukmendapatkan jawapan terhadap input kueri daripada pengguna dengan pangkatan dokumen yang relevan mengikut nilai persamaan antara dokumen jawapan dan set soalan dalam pangkalan data indeks.

Pengukuran jarak vektor antara kueri dan dokumen digunakan untuk memangkat dokumen yang dicapai. Dokumen maklumat dan set soalan Saie disimpan sebagai beg perkataan korpus di dalam pangkalan data leksikel bagi mengurangkan kekompleksan dokumen dan memudahkan pemprosesan data.

PEMBERAT PERKATAAN

Setiap perkataan bagi semua maklumat korpus jawapan dan set soalan akan dibuat pengiraan nilai pemberat. Pengiraan nilai pemberat (TFIDF) bertujuan untuk mendapatkan keseimbangan antara soalan dan jawapan. Pengumpukan nilai pemberat kepada setiap perkataan untuk mendapatkan kata kunci sesuatu dokumen. Bagi pemberat perkataan yang tinggi dalam sesuatu dokumen dapat menentukan tahap kerelevanan sesuatu dokumen.

$$\text{Pemberat} = tf_{i,d} * \log_{10}\left[\frac{N+1}{n_i}\right]$$

Di mana,

a) $tf_{i,d}$ = nilai kekerapan sesuatu perkataan *i* muncul di dalam dokumen *d*
b) N = jumlah dokumen di dalam koleksi pangkalan data korpus Saie
c) Nilai 1 = mewakili satu soalan bagi setiap pengiraan nilai pemberat
d) n_i = bilangan dokumen dalam koleksi yang mengandungi perkataan *i*

PERSAMAAN PADANAN

Ukuran padanan untuk mendapatkan tahap padanan antara dua vektor iaitu soalan dan jawapan. Dengan nilai padanan antara soalan dan jawapan membolehkan sistem ini memangkat dokumen yang dicapai mengikut susunan yang teratas adalah dokumen yang paling penting sehingga ke bawah nilai yang paling rendah. Tiga jenis persamaan padanan iaitu *Inner Product,* Kosinus dan *Jaccard Coefficient.*

$$\text{InnerSim}(Di, Q) = \sum_{k=1}^{t} (dik \,.\, qk)$$

$$\text{CosSim}(Di, Q) = \frac{\sum_{k=1}^{t} (dik \,.\, qk)}{\sum_{k=1}^{t} d_{i\,t} \cdot \sum_{k=1}^{t} qk}$$

$$\text{JaccardSim}(Di, Q) = \frac{\sum_{k=1}^{t} (dik \,.\, qk)}{\sum_{k=1}^{t} d_{ik}^{2} + \sum_{k=1}^{t} q_k^{2} - \sum_{k=1}^{t} (dik \,.\, qk)}$$

Dimana,

a) d_{ik} = Pemberat TFIDF perkataan *k* dalam dokumen *i*
b) q_k = Pemberat TFIDF perkataan *k* dalam kueri

KEJITUAN, DAPATAN DAN F-MEASURE

Alat asas dalam proses mengukur prestasi pemangkatan dokumen capaian adalah dengan menggunakan pengukuran kejituan dan dapatan. Bahagian capaian maklumat relevan diukur sebagai kejituan, manakala dapatan pula adalah bahagian dokumen yang berkemungkinan relevan dengan maklumat yang dicari. Formula bagi pengukuran kejituan dan dapatan [2]:

$$\text{Dapatan(R)} = \frac{\textit{Bilangan dokumen relevan dicapai sistem}}{\textit{Jumlah dokumen relevan dicapai sistem}}$$

$$\text{Kejituan (P)} = \frac{\textit{Bilangan dokumen relevan dicapai sistem}}{\textit{Bilangan dokumen dicapai system}}$$

$$\text{F-measure} = 2 * \frac{PR}{P+R}$$

DAPATAN KAJIAN

Eksperimen dijalankan melibatkan dua pengujian iaitu Ujian Tertutup dan Ujian terbuka. Penilaian Ujian Tertutup terhadap QA Ibadah Saie mengandungi 20 set jawapan dan 15 set soalan. Manakala, Ujian Terbuka dilakukan berdasarkan 5 soalan yang diambil daripada forum dan diuji ke atas 15 pasangan set soalan dan jawapan.

1. Ujian Tertutup

Jenis Penilaian	Persamaan Kosinus	Inner Product	Jaccard Coeficient
Kejituan	80%	53%	40%
Dapatan	86%	73%	60%
F-measure	83%	61%	48%

2. Ujian Terbuka

Jenis Penilaian	Persamaan Kosinus	Inner Product	Jaccard Coeficient
Kejituan	60%	20%	40%
Dapatan	75%	50%	50%
F-measure	67%	29%	44%

PERLUASAN KAJIAN

1. Menyediakan satu alternative perkataan seperti thesaurus untuk perkataan yang berkaitan dengan topic Haji dan Umrah khususnya ibadah Saie.
2. Menyedari kekurangan sumber digital bagi topik Haji dan Umrah khususnya ibadah Saie membuka banyak ruang untuk pembangunan aplikasi pengindeksan maklumat berbentuk digital.
3. Mengembangkan maklumat Ialam tentang Haji dan Umrah dalam Pangkalan Data Pengetahuan khususnya Wordnet dapat meningkatkan prestasi persamaan metric dalam pencarian kesepadanan kueri.

KESIMPULAN

Keputusan yang dicapai membuktikan kaedah Model Ruang Vektor menggunakan nilai pemberat dan padanan kosinus boleh dilaksanakan dan sistem ini boleh diaplikasikan ke dalam QA ibadah Saie berdasarkan hasil keputusan ketiga-tiga penilaian boleh diklasifikasikan sebagai seimbang dan boleh dipercayai.

RUJUKAN

[1] Christoper D. Manning, Prabhakar Raghavan dan Hinrich Schutze. 2008. Introduction to Information Retrieval. Cambridge: Cambridge University Press.

[2] Jurafsky, D. & Martin, J. H., 2000. Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics And Speech Recognition. New Jersey: Prentice Hall.

[3] Gracinda Carvalho, David Martins de Martos & Vitor Rotio., 2007. Document Retrieval For Question Answering: A Quantitative Evaluation of Text Preprocessing. ACM.

[4] Zhengtao Yu, Huanyun Zong, Yangbo Xu, Jianyi Guo, Yu Mao dan Xiangyan Meng., 2009. FAQ Extracting and Domain Filtering Based on Improved Bayes. International Conference on Web Information Sistems and Mining on IEEE.