

## HEURISTICS-BASED METHOD FOR HEAD AND MODIFIER DETECTION IN MALAY SENTENCES FROM THE CULTURAL HERITAGE DOMAIN

SUHAIMI AB RAHMAN  
NAZLIA OMAR

### ABSTRAK

*The process of detection for the head and modifier in Malay sentences from the cultural heritage domain is difficult to identify. This is due to the position of head and modifier which varies in sentences depending on the sentence structures. Hence, there are different point of views about the theory and concept of detection for the head and modifier in a compound noun that have been discussed by language experts. Additionally, the existing research is also limited especially in the areas of computational linguistics. Therefore, research should be conducted to identify appropriate methods especially used in the detection of head and modifier which appear in Malay sentences from the cultural heritage domain. The aim of this study is to construct a list of heuristic rules to be used for detecting the position of compound nouns in Malay sentences from cultural heritage domain. By using 15 rules, the position of head and modifier that exist in a compound noun can also be detected. These rules are called heuristic rules. The purpose of formulating these 15 rules is to detect the head and modifier that exist in the Malay sentences from the cultural heritage domain. To measure the accuracy of the results, precision, recall and F1-score values are used. Based on the results of the experiments, Sentence Structure of Malay Cultural Heritage Domain (SADWBM) have an F1-score of 80.4% compared to Noun Phrase Structure (SFN) which is 56%. Consequently, SADWBM shows better scores compared to SFN. Therefore it is clear that the approach used in this study is effective in resolving the identified problems.*

*Keywords: Head and Modifier, Noun Phrase Structure (SFN), Sentence Structure of Malay Cultural Heritage Domain (SADWBM), Heuristic Rule.*

## KAEDAH HEURISTIK BAGI MENGECEM INTI DAN PENERANG DALAM AYAT BAHASA MELAYU DARI DOMAIN WARISAN BUDAYA

### ABSTRAK

Proses pengecaman bagi mengenal pasti kedudukan inti dan penerang yang hadir dalam ayat dari domain warisan budaya Melayu adalah sukar. Ini kerana kedudukan inti ayat adalah berbeza mengikut struktur ayat masing-masing. Pelbagai pandangan tentang teori dan konsep pengecaman inti dan penerang dalam kata majmuk oleh pakar bahasa dikupas. Selain daripada itu, kajian sedia ada juga tidak membincang secara terperinci mengenai pengecaman inti dan penerang dalam bentuk linguistik komputer. Justeru, penyelidikan perlu dilakukan bagi mengenal pasti kaedah yang sesuai diguna khususnya dalam mengecam inti dan penerang dalam ayat dari domain warisan budaya Melayu. Kajian ini bertujuan membina senarai petua heuristik yang diguna bagi mengecam kedudukan kata majmuk dalam ayat dari domain warisan budaya Melayu. Menggunakan petua heuristik membolehkan kedudukan inti dan penerang dalam kata majmuk dikenal pasti dengan mudah. Sebanyak 15 jenis petua dibina yang berfungsi mengecam inti dan penerang yang wujud dalam ayat tersebut. Unit penilaian yang diguna bagi mengukur ketepatan bagi setiap uji kaji ini ialah kejituan, dapatan semula dan skor-F1. Berdasarkan keputusan uji kaji Struktur Ayat Dari Domain Warisan Budaya Melayu (SADWBM) mempunyai nilai kejituan sebanyak 80.4% dan berbeza dengan Struktur Frasa Nama (SFN), nilai kejituan ialah sebanyak 56%. Oleh yang demikian SADWBM menunjuk ketepatan yang baik berbanding dengan SFN dan pada masa yang sama membuktikan bahawa pendekatan yang diguna adalah baik dalam menyelesaikan masalah mengecam inti dan penerang dalam ayat dari domain warisan budaya Melayu.

Kata Kunci: Inti dan Penerang, Struktur Frasa Nama (SFN), Struktur Ayat Daripada Domain Warisan Budaya Melayu (SADWBM), Petua Heuristik.

## PENGENALAN

Kajian ini merupakan salah satu daripada kajian dalam bidang warisan budaya linguistik Melayu yang memberi fokus kepada bidang sintaksis. Dalam bidang sintaksis, kajian menjurus ke arah binaan kategori frasa nama dan binaan kata nama majmuk yang terkandung inti dan penerang di dalamnya. Dalam setiap frasa terdapat perkataan yang berfungsi sebagai inti frasa. Inti frasa ialah perkataan yang penting berbanding dengan perkataan lain. Perkataan lain tersebut dikenali sebagai penerang (Abdullah, 2004). Misalnya dalam ayat “*Budaya melambang identiti sesebuah negara*”. Inti ayat di atas ialah *budaya*, manakala perkataan yang lain seperti *melambang*, *identiti*, *sesebuah* dan *negara* adalah merupakan penerang kepada perkataan inti iaitu *budaya*. Dalam bahasa Melayu terdapat empat jenis frasa iaitu frasa nama, frasa adjektif, frasa kerja dan frasa sendi nama. Dalam setiap jenis frasa tersebut terdapat intinya masing-masing. Contohnya, frasa nama mengandungi inti frasa nama dan frasa adjektif mempunyai inti frasa adjektif (Abdullah, 2004, 2009). Sebagai contoh ayat *warisan Melayu*. Perkataan *warisan* adalah dari kelas kata nama dalam bentuk inti kepada penerang perkataan *Melayu* daripada kelas kata nama.

Menurut Abdullah (2009) dan Nik Safiah et al. (2010), inti dan penerang hadir bersama dalam suatu binaan ayat yang sempurna. Oleh yang demikian, satu kaedah bagi mengecam kata majmuk bersama inti dan penerang yang wujud dalam sesuatu frasa diperlu. Kajian ini menggunakan pendekatan yang berupa rumus dan petua heuristik. Rumus adalah susunan kelas kata yang relevan dan logikal bagi membentuk satu rentetan ayat yang sempurna manakala petua heuristik bermaksud satu pernyataan bersyarat yang menyemak setiap syarat sebelum sesuatu keputusan yang logikal dan benar dihasil (Negnevitsky, 2005).

## LATAR BELAKANG

Kajian memfokus kepada struktur ayat daripada domain warisan budaya Melayu, binaan kata majmuk bersama inti dan penerang dan peranan petua sebagai representasi pengetahuan.

## WARISAN BUDAYA MELAYU

Warisan budaya mencakupi amalan dan cara hidup pelbagai kaum dalam sesebuah negara. Hal ini termasuklah pakaian, makanan, adat resam dan seni. Tamadun sesuatu bangsa selalunya dapat dilihat pada warisan seni budayanya. Justeru, satu usaha perlu dilakukan bagi mengekal elemen budaya sebagai satu warisan untuk tatapan generasi akan datang. Amalan budaya tradisional di Malaysia melambang sifat majmuk masyarakat di negara ini yang berbilang kaum (Mazlan, 2008). Bagi menerap pemahaman unsur warisan budaya yang bersifat majmuk, maka satu kaedah menterjemah sumber asal seperti daripada bahasa Melayu kepada bahasa Inggeris, Bahasa Tamil dan Bahasa China perlu diusahakan. Satu langkah awal yang dikenal pasti ialah membuat kajian ke atas struktur binaan ayat khususnya daripada domain warisan budaya Melayu. Pemahaman kedudukan susun atur perkataan dalam bentuk inti dan penerang adalah penting bagi menentu jenis gabungan dua frasa daripada satu binaan frasa atau ayat yang sempurna. Selain daripada itu, kajian terhadap kedudukan inti dan penerang juga boleh membantu memahami makna sesuatu ayat. Ini secara tidak langsung dapat membantu proses penterjemahan yang tepat dan berkualiti, khususnya dari segi makna ayat yang diterjemah.

## BINAAN KATA MAJMUH, INTI DAN PENERANG

Bisetto dan Scalise (2005), Fan et al. (2003) dan Guevara (2013) mendefinisikan bahawa kata majmuk dibina daripada sekurang-kurangnya dua perkataan. Perkataan ini wujud daripada kombinasi antara beberapa perkataan daripada perkataan yang lain. Misalnya jenis kata majmuk dalam bahasa Inggeris boleh dilihat seperti contoh berikut (Brants, 2003; Liddy, 2001; Rus et al., 2002):

- i. *sauce (noun) + pan (noun) = saucepan (periuk)*
- ii. *soft (adjective) + ware (noun) = software (perisian)*
- iii. *break (verb) + fast (noun) = breakfast (sarapan)*
- iv. *washing (verbal) + machine (noun) = washing machine (mesin basuh)*
- v. *sun (noun) + rise (verb) = sunrise (matahari terbit)*
- vi. *in (preposition) + flux (noun) = influx (kemasukan)*
- vii. *out (preposition) + put (verb) = output (hasil)*
- viii. *check (verb) + out (preposition) = checkout (keluar)*

Walaupun terdapat kekurangan yang boleh mempengaruhi proses menentu susun atur sesuatu binaan kata majmuk, namun kajian yang komprehensif mengenai kata majmuk perlu dijalankan dari semasa ke semasa. Tiga faktor kekurangan yang dikenal pasti ialah pertama, pemahaman makna kata majmuk dalam konteks susunan perkataan yang perlu berada dalamnya (Abdullah, 2004; Nik Safiah et al., 2010; Tratz & Hovy, 2010). Kedua, masalah dari segi struktur bagaimana perkataan dihubungkan untuk membentuk ayat, khususnya dalam bentuk domain bahasa yang berbeza (Bisetto & Scalise, 2005; Ditters & Koster, 2004) dan ketiga ialah jenis pendekatan yang digunakan bagi mengenal pasti struktur binaan kata majmuk bagi tujuan penyelesaian dalam bentuk linguistik komputer (Abdullah, 2004; Bisetto & Scalise, 2005; Dien, 2002; Fan et al., 2003; Nik Safiah et al., 2010). Oleh yang demikian, kajian yang dilakukan ini adalah untuk mengenalpasti permasalahan yang timbul serta penyelesaian yang berkesan dalam menyelesaikan binaan kata majmuk khususnya dalam bahasa Melayu.

Dalam binaan kata majmuk untuk bahasa Melayu, kaedah kategori digunakan. Faridah (2014) dan Nik Safiah et al. (2010) membahagi dua konstituen frasa nama kepada dua jenis iaitu i) inti dan penerang, dan ii) inti dan inti. Kajian seperti ini merupakan kajian yang komprehensif dan digunakan sebagai panduan untuk membina kata majmuk dalam bahasa Melayu (Faridah, 2014; Nik Safiah et al., 2010). Oleh itu, setiap pasangan perkataan dalam kata majmuk dalam bahasa Melayu dikategorikan kepada 19 jenis kumpulan yang berbeza. Contohnya ialah 13 jenis kumpulan daripada inti dan penerang nama, dan enam jenis kumpulan daripada inti dan penerang bukan nama.

## PETUA SEBAGAI REPRESENTASI PENGETAHUAN

Dalam bidang linguistik komputer, penterjemahan daripada konsep dan teori struktur tata bahasa kepada pengaturcaraan komputer adalah perkara penting yang harus dikaji. Justeru, pendekatan rumus dan petua heuristik dibina bagi menyokong keperluan pengaturcaraan komputer (Zulkifley et al., 2007). Daripada rumus ini, petua diformulasi untuk memenuhi syarat bagi setiap rumus yang dibina. Negnevitsky (2005) berpendapat petua adalah satu daripada teknik asas yang digunakan untuk mewakili pengetahuan bagi setiap sistem yang berasaskan kepada pemrosesan logik. Sehubungan itu, Negnevitsky (2005) menyatakan untuk sesuatu sistem yang berasaskan kepada penggunaan petua, ia boleh dikelas kepada beberapa jenis petua iaitu i) hubungan ii) cadangan iii) arahan iv) strategi, dan v) heuristik. Contoh penggunaan setiap jenis petua seperti berikut:

i) Petua hubungan: Petua ini adalah yang paling asas yang menghubungkan di antara syarat dan keputusan. Contohnya:

JIKA (syarat1: *tangki minyak kosong*) MAKA (keputusan: *enjin kereta tidak hidup*)

ii) Petua cadangan: Petua ini menyemak beberapa syarat unsur sebelum mencadang sesuatu keputusan. Contohnya:

JIKA (syarat1: *musim luruh tiba*) DAN (syarat2: *langit berawam*) DAN (syarat3: *ramalan hujan renyai-renyai*) MAKA (keputusan: *dinasihatkan sediakan payung*)

iii) Petua arahan: Petua ini menyemak beberapa syarat dan keputusan yang berbentuk arahan. Contohnya:

JIKA (syarat1: *enjin kereta tidak hidup*) DAN (syarat2: *tangki minyak kosong*)  
MAKA (keputusan: *diarahkan supaya isi minyak kereta*)

iv) Petua strategi: Petua ini menyemak syarat yang perlu dipenuhi sebelum keputusan dalam bentuk strategi dijana. Petua ini sedikit berbeza berbanding petua lain, kerana melibatkan beberapa langkah susulan yang perlu diambil bagi melengkap proses yang menghasilkan keputusan. Contohnya:

JIKA (syarat1: *enjin kereta tidak hidup*)  
MAKA (keputusan: *periksa tangki minyak*); Langkah 1 selesai.  
JIKA (syarat2: *langkah 1 selesai*) AND (syarat2: *tangki minyak penuh*)  
MAKA (keputusan: *periksa bateri kereta*); Langkah 2 selesai

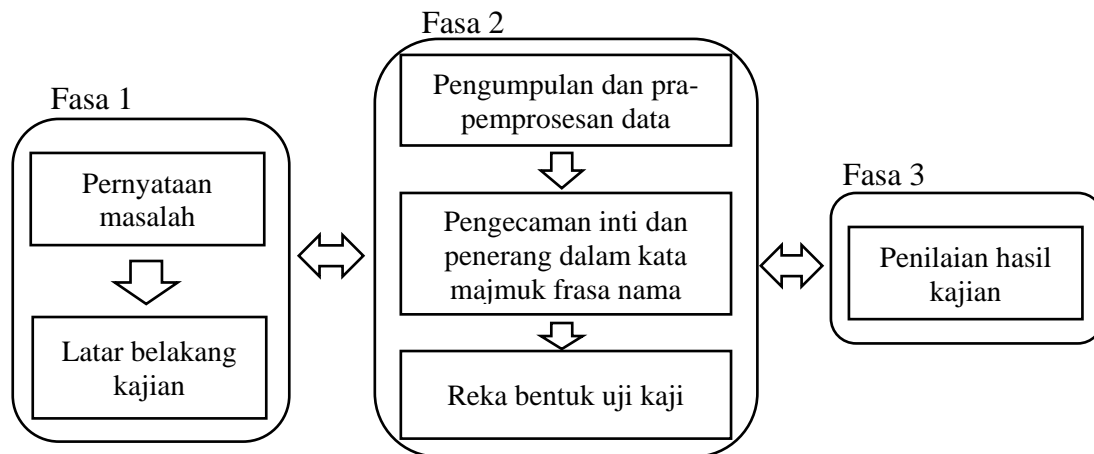
v) Petua heuristik: Petua ini menyemak beberapa syarat yang hendak dinilai dan keputusan dihasil sekiranya semua syarat mengembali nilai benar. Petua ini bersifat jujukan dalam membuat sesuatu keputusan terhadap nilai pembolehubah yang perlu disemak. Setiap pembolehubah dibanding dengan nilai pemalar. Apabila semua pembolehubah dipenuhi maka petua mencadang keputusan yang relevan bagi pembolehubah tersebut. Contohnya:

JIKA (syarat 1: *tumpahan ialah cecair*) DAN (syarat 2: *nilai PH tumpahan < 6*) DAN (syarat 3: *tumpahan berbau masam*) MAKA (keputusan: *tumpahan adalah jenis asid asetik*)

Pemilihan jenis petua yang sesuai adalah bergantung kepada jenis keperluan kajian yang dijalankan. Setiap petua berperanan menyelesaikan masalah mengikut keperluan spesifikasi kajian yang berbeza. Oleh itu, penggunaan petua dianggap teknik yang berkesan untuk dieksploitasi bagi menghasilkan penyelesaian yang berkesan dalam membuat sesuatu keputusan (Abdullah et al., 2008; Gerald & Zuraidah, 2006; Phang & Zarina, 2012).

## REKA BENTUK KAJIAN

Reka bentuk kajian adalah penting dalam menggambar aktiviti penyelidikan yang menyeluruh yang dilakukan sepanjang kajian. Reka bentuk kajian terbahagi kepada tiga fasa iaitu Fasa 1, Fasa 2 dan Fasa 3. Semua modul yang terdapat dalam setiap fasa adalah seperti yang ditunjuk dalam Rajah 1.



RAJAH 1. Reka bentuk kajian

### 1. Fasa 1: Pemahaman Masalah Kajian

Fasa 1 adalah merupakan fasa terpenting dalam peringkat awal kajian. Fasa ini menyumbang ke arah pemahaman tentang sesuatu permasalahan kajian yang timbul. Selanjutnya, latar belakang kajian dijalankan secara komprehensif bagi mengkaji, menilai, mengkritik teknik, proses dan pendekatan yang berkaitan berdasarkan kepada kajian sedia ada.

### 2. Fasa 2: Kerangka Aliran Proses

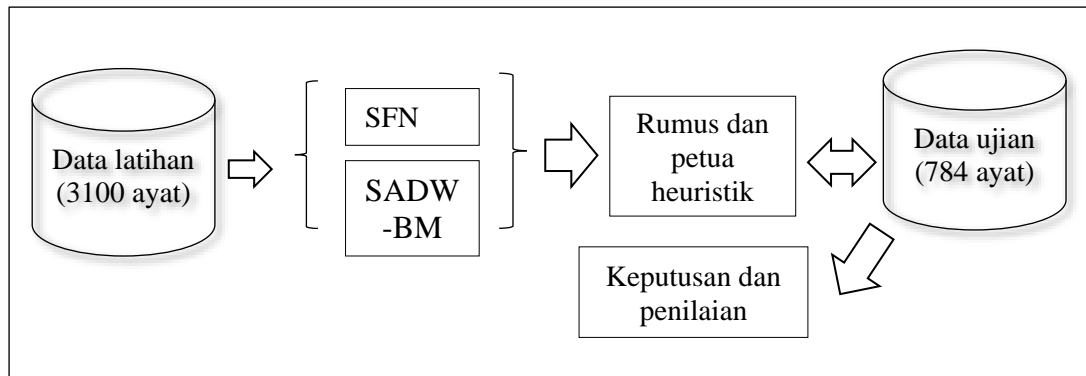
Fasa ini penting bagi memasti jenis aktiviti kajian yang dilakukan bagi tujuan mencapai objektif. Fasa ini melibatkan tiga modul iaitu pengumpulan dan pra-pemprosesan data, pengecaman inti dan penerang dalam kata majmuk frasa nama bahasa Melayu, dan reka bentuk uji kaji.

### 3. Fasa 3: Penilaian

Fasa ini menerangkan teknik yang diguna untuk menilai hasil kajian yang berupa kata majmuk yang di dalamnya terkandung inti dan penerang. Setiap hasil dalam bentuk kata majmuk disemak sama ada relevan atau sebaliknya dengan berpandu kepada kepada dua kategori kata majmuk iaitu i) inti dan penerang nama dan ii) inti dan penerang bukan nama yang dikaji dan ditakrif oleh pakar Bahasa seperti Abdullah (2007), Nik Safiah et al. (2010) dan Faridah (2014)

## REKA BENTUK DATA KAJIAN

Reka bentuk data kajian memperlihatkan dua jenis kategori data yang berbeza yang diguna dalam kajian. Kedua-dua jenis kategori data adalah seperti yang ditunjuk dalam Rajah 2.



RAJAH 2. Reka bentuk data kajian

Rajah 2 menunjukkan dua jenis set data yang diguna iaitu set data latihan dan set data ujian. Pengumpulan set data latihan adalah bertujuan membina rumus serta petua heuristik yang berpandu kepada SFN. Selain daripada itu, data ini juga diguna bagi tujuan membangun struktur kerangka yang diguna untuk membina SADWBM. Manakala set data ujian pula diguna sebagai data ujian dalam setiap uji kaji yang dinyatakan. Dari segi pembahagian ayat, sejumlah 3100 ayat dikelompok dalam set data latihan, manakala 784 ayat pula diletak dalam set data ujian.

## SENARIO PENGUJIAN

Senario pengujian dilakukan bagi memasti proses yang berlaku pada setiap uji kaji boleh dilakukan dengan tersusun. Tujuan uji kaji ini dijalankan ialah bagi mendapatkan purata dan peratus nilai kejituan, dapatan semula dan skor-F1 di antara dua kerangka iaitu SNF dan SADWBM. Secara umumnya, ketepatan hasil bagi setiap kerangka adalah bergantung kepada jumlah petua heuristik yang dibina bagi menyemak inti dan penerang yang terdapat dalam ayat daripada domain warisan budaya Melayu.

Empat langkah yang dikenal pasti bagi menjalankan uji kaji yang dinyatakan adalah seperti berikut:

Langkah 1: Data ujian terdiri daripada 784 ayat yang diguna sebagai data uji kaji.

Langkah 2: Dapatkan nilai purata dan peratus bagi setiap nilai kejituan, dapatan semula dan skor-F1 bagi setiap jenis hubungan rantaian inti penerang ayat untuk SADWBM.

Langkah 3: Membanding hasil keputusan daripada SFN dan SADWBM.

Langkah 4: Bina kesimpulan daripada hasil perbandingan tersebut.

## PERBINCANGAN

Berdasarkan uji kaji yang dilakukan, perbincangan yang menyeluruh dilakukan bagi merumus sama ada pendekatan yang diguna dalam kajian adalah baik atau sebaliknya dalam menangani masalah seperti yang dikenal pasti. Contoh keputusan yang dikompil daripada semua nilai

kejitian, dapatan semula dan skor-F1 untuk SADWBM yang diperoleh daripada uji kaji dengan nilai kejitian, dapatan semula dan skor-F1 daripada SFN adalah seperti berikut:

#### ANALISIS SFN

Analisis SFN untuk nilai kejitian, dapatan semula dan skor-F1 dilakukan dengan mengguna sejumlah 50 ayat yang dipilih secara rawak daripada 784 ayat set data ujian, dan kemudian dianalisis. Kiraan nilai kejitian, dapatan semula dan skor-F1 yang diperoleh daripada analisis 50 ayat untuk struktur frasa nama adalah seperti berikut:

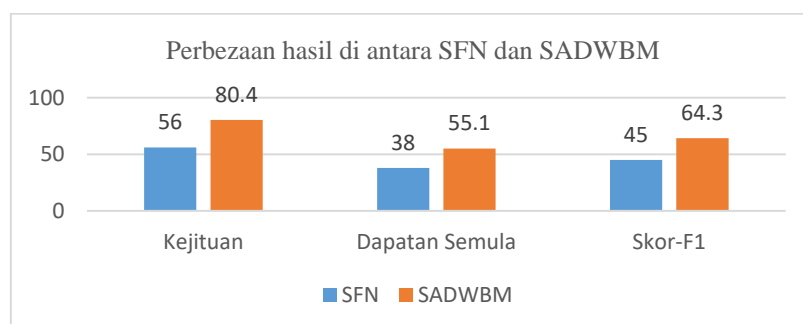
JADUAL 1. Kiraan ketepatan SFN

Pembolehubah	Nilai
R = Jumlah kata majmuk yang relevan berjaya dicapai dari T	18
TDU = Jumlah set data ujian	50
T = Jumlah kata majmuk dalam setiap ayat berkaitan	26
RM = Jumlah kata majmuk relevan yang disemak secara manual dari R	10
TRM = Jumlah kata majmuk tidak relevan yang disemak secara manual dari R	8
TRB = Jumlah kata majmuk yang tidak relevan berjaya dicapai dari T	16
Ketepatan = $(RM / (RM + TRM))$	0.56
Dapatan Semula = $(RM / (RM + TRB))$	0.38
Skor-F1 = $2 * ((K * DS) / (K + DS))$	0.45

Keputusan yang diperoleh, menunjukkan nilai bagi setiap kejitian, dapatan semula dan skor-F1 adalah sebanyak 56%, 38% dan 45%. Peratus nilai kejitian sebanyak 56% adalah tinggi sedikit daripada peratus nilai dapatan semula. Ini menepati setiap set data ujian yang diuji serta mempunyai ketepatan ujian yang agak tinggi. Kejitian merujuk kepada ketepatan penentuan rumus yang terdapat dalam struktur frasa nama pada set data ujian. Secara umum, peratusan keseluruhan masih berada pada kedudukan yang agak tinggi iaitu sebanyak 45% yang merujuk kepada skor-F1. Pengukuran penilaian pada unit skor-F1 ini adalah hasil akhir bagi kedua-dua unit penilaian dapatan semula dan kejitian. Dapat disimpulkan bahawa sebanyak 45% dijana oleh skor-F1 yang menunjukkan peranan rumus dalam struktur frasa nama adalah baik dan tepat dalam menentu kedudukan kata majmuk yang hadir dalam sesuatu ayat yang berbentuk frasa nama.

#### ANALISIS SFN DAN SADWBM

Perbincangan yang menyeluruh dilakukan bagi merumuskan sama ada pendekatan yang diguna dalam kajian adalah baik atau sebaliknya dalam menangani masalah seperti yang dikenal pasti dalam kajian. Contoh keputusan yang dikompil daripada semua nilai kejitian, dapatan semula dan skor-F1 dari SFN dan SADWBM adalah seperti berikut:



RAJAH 3. Perbezaan Hasil SFN dan SADWBM

Berdasarkan keputusan uji kaji yang ditunjukkan dalam Rajah 3, SADWBM mempunyai nilai kejituan sebanyak 80.4% berbanding dengan nilai kejituan SFN sebanyak 56%. Ini menepati set data ujian yang diuji dan merujuk kepada ketepatan peranan petua heuristik untuk mengecam inti dan penerang dalam kata majmuk frasa nama pada set data ujian. Bagaimanapun, peratus keseluruhan SADWBM masih berada pada kedudukan yang agak tinggi iaitu sebanyak 45% pada SFN dan 64.3% pada SADWBM yang merujuk kepada skor-F1. Ini bererti, SADWBM menunjukkan ketepatan yang baik berbanding dengan SFN dan pada masa yang sama membuktikan pendekatan yang diguna dalam kajian adalah baik dalam menyelesaikan masalah mengecam inti dan penerang dalam ayat dari domain warisan budaya Melayu.

## KESIMPULAN

Kajian ini membina senarai petua heuristik yang diguna bagi mengecam kedudukan kata majmuk yang hadir dalam ayat dari domain warisan budaya Melayu. Terdapat tiga proses yang dilaksana iaitu. Pertama, melakukan analisis ayat bahasa Melayu bagi membina SADWBM. Kerangka ini dijadikan panduan bagi membina rumus yang berkaitan. Kedua, setelah semua rumus ditakrif, rumus tersebut kemudiannya diimplementasi ke dalam bentuk linguistik komputer dengan mengguna pendekatan petua heuristik. Ketiga, mengira nilai ketepatan uji kaji. Bagi menilai ketepatan hasil kajian, uji kaji terhadap pengukuran ketepatan SFN dan SADWBM dilakukan. Hasil uji kaji menunjukkan pendekatan yang diguna bagi menyelesaikan masalah mengecam inti dan penerang dalam ayat dari domain warisan budaya Melayu adalah baik. Hasil kajian dapat membantu para penyelidik dan pakar dalam bidang pemprosesan bahasa tabii, khususnya dalam membangun aplikasi penghurai sintaksis bahasa, penterjemahan bahasa, hierarki ontologi dan kajian semantik khususnya yang berkaitan dengan sumber ayat dari domain warisan budaya Melayu.

## RUJUKAN

- Abdullah Hassan. 2004. *Tatabahasa bahasa Melayu, morfologi dan sintaksis*. Bentong: PTS Publications and Distributors.
- Abdullah Hassan. 2009. *Tatabahasa pedagogi untuk sekolah menengah*. Kuala Lumpur: PTS Professional Publishing.
- Bisetto, A. & Scalise, S. 2005. The classification of compounds. *Lingue e linguaggio*, 4(2): 319 -320.
- Brants T, Google Inc .2003. Natural Language Processing in Information Retrieval, *CLIN 2003*:1-13.
- Dien, D. 2002. Cognitive linguistics approach to Vietnamese noun compounds. *Mon-Khmer Studies: Journal of Southeast Asian Linguistics and Languages*, 32:145–161.
- Ditters, W. E. & Koster, C. H. 2004. Transducing Arabic phrases into Head-Modifier (HM) pairs for Arabic information retrieval. *Proceedings of the NEMLAR 2004 International Conference on Arabic Language Resources and Tools*. Paris: ELDA, 148–152.
- Guevara, E. 2013. *Compound Words. The Encyclopedia of Applied Linguistics*. Oxford: Blackwell Publishing Ltd.
- Fan, J., Barker, K. & Porter, B. 2003. The knowledge required to interpret noun compounds, *IJCAI*, 1483–1485.
- Faridah Nasir. 2014. *Sintaksis Bahasa Melayu*. Puchong: Penerbitan Multimedia.
- Gerald O.K. & Zuraidah Mohd Don. 2006. *Word class in Malay – a corpus-based approach*. 1<sup>st</sup> Ed. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- Liddy, E.D. 2001. Natural Language Processing. In: *Encyclopedia of Library and Information Science*, 2<sup>nd</sup> Ed. New York: Marcel Decker.
- Mazlan. 2008. Kepentingan budaya tradisonal kita. <https://mazlan66.wordpress.com/2008/03/16/kepentingan-budaya-tradisonal-kita/> [16 Mac 2008]
- Negnevitsky, M. 2005. *Artificial Intelligence*. 2<sup>nd</sup> ed. London: Pearson Education Limited.



- Nik Safiah Karim, Farid M. Onn, Hashim Haji Musa & Abdul Hamid Mahmood. 2010. *Tatabahasa Dewan*. Edisi Ketiga. Ampang: Dawama Sdn. Bhd.
- Phang, See Wang & Zarina Shukur. 2012. *Asas Teknik Pemprosesan Bahasa*. Cetakan Pertama. Bangi: Universiti Kebangsaan Malaysia.
- Rus, V., Moldovan, D. I. & Bolohan, O. 2002. Bracketing Compound Nouns for Logic Form Derivation. *FLAIRS Conference*, 198–202.
- Tratz, S. & Hovy, E. 2010. A taxonomy, dataset, and classifier for automatic noun compound interpretation. *Proceedings of the 48<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*, 678–687. Uppsala, Sweden: Association for Computational Linguistics, 678–687.
- Zulkifley Hamid, Ramli Md Salleh & Rahim Aman. 2007. *Linguistik Melayu*. Edisi Kedua. Bangi: Penerbit Universiti Kebangsaan Malaysia.

Suhaimi Ab Rahman  
Jabatan Kejuruteraan Perisian,  
Kolej Komputer Sains dan Teknologi Maklumat,  
Universiti Tenaga Nasional (UNITEN),  
km7, Jalan Ikaram-Uniten, 43009 Kajang, Selangor, Malaysia.  
smie@uniten.edu.my

Nazlia Omar  
Fakulti Teknologi dan Sains Maklumat (FTSM),  
UKM, Bangi, 43600, Selangor, Malaysia.  
nazlia@ukm.edu.my

Received: 30 May 2016  
Accepted: 14 March 2017