



## INTERPRETASI PERIBAHASA INDONESIA DENGAN METODE *PATTERN MATCHING* BERBASIS KATA KUNCI

**Farida Yunita<sup>1)</sup>, Riska Dwi Handayani<sup>2)</sup>, Moch Ali Mahmudi<sup>3)</sup>**

<sup>1,2,3)</sup>“Manajemen Informatika” STMIK BINA PATRIA Magelang  
Email : frida\_diajeng@stmikbinapatria.ac.id<sup>1)</sup>, riska@stmikbinapatria.ac.id<sup>2)</sup>,  
ali@stmikbinapatria.ac.id<sup>2,3)</sup>

### Abstract

*Indonesian proverbs are an inseparable part of Indonesian cultural heritage that has been rooted in the daily lives of people from various tribes and regions. In the era of globalization and digitalization, there is a gap between technological progress and cultural preservation, resulting in a decreasing understanding of Indonesian proverbs among the public, which in turn can lead to the erosion of the nation's cultural heritage. One solution to this problem is a Keyword-based Pattern Matching approach for matching keywords in each proverb. Pattern Matching has the ability to identify specific lexical patterns, efficiency in mapping keywords to contextual meanings, flexibility in handling linguistic variations and suitability with the characteristics of proverbs that have core keywords, ease in representing expert knowledge into digital systems. The purpose of this study is to develop a matching algorithm, namely Jaccard similarity, for matching keywords for each proverb so that the proverb can be interpreted. Jaccard similarity has three stages, namely: keyword input, prepossessing and pattern matching. The results of this study from 30 scenarios from four keyword categories, the accuracy of matching keywords with proverb keywords average is 75%, so that interpretations of Indonesian proverbs can be generated.*

**Keywords :** Expert System, Proverbs, Interpretation, Pattern Matching, Jaccard Similarity

### Abstrak

Peribahasa Indonesia merupakan bagian tak terpisahkan dari warisan kebudayaan Indonesia yang telah mengakar dalam kehidupan sehari-hari masyarakat dari berbagai suku dan daerah. Dalam era globalisasi dan digitalisasi, terdapat kesenjangan antara kemajuan teknologi dan pelestarian budaya akibatnya pemahaman terhadap peribahasa Indonesia semakin menurun di kalangan masyarakat, yang pada akhirnya dapat mengakibatkan terkikisnya warisan budaya bangsa. Salah satu solusi permasalahan adalah dengan pendekatan *Pattern Matching* berbasis Kata Kunci. Kata Kunci untuk mencari dan mencocokan kata-kata kunci tertentu yang ada di dalam peribahasa. *Pattern Matching* memiliki kemampuan mengidentifikasi pola leksikal spesifik, efisiensi dalam memetakan kata kunci ke makna kontekstual, fleksibilitas menangani variasi *linguistic* dan kesesuaian dengan karakteristik peribahasa yang memiliki kata kunci inti, kemudahan dalam merepresentasikan pengetahuan pakar ke dalam sistem digital. Tujuan penelitian ini adalah mengembangkan algoritma *matching* yaitu *jaccard similarity* untuk pencocokan kata kunci setiap peribahasa sehingga peribahsa dapat di *interpretasi*. *Jaccard similarity* memiliki tiga tahap, yaitu: *input* kata kunci, *prepossessing* dan *pattern matching*. Hasil penelitian ini dari 30 skenario dari empat kategori kata kunci, akurasi kecocokan kata kunci dengan kata kunci peribahasa rata-rata 75%, sehingga interpretasi peribahasa Indonesia dapat dihasilkan.

**Kata kunci:** Sistem Pakar, Peribahasa, *Interpretasi*, *PatternMatching*, *Jaccard Similarity*.

### 1. Pendahuluan

Peribahasa Indonesia merupakan bagian tak terpisahkan dari warisan kebudayaan Indonesia yang telah mengakar dalam kehidupan sehari-hari masyarakat dari berbagai suku dan daerah. Peribahasa tidak hanya mencerminkan identitas bangsa, tetapi juga berfungsi untuk memahami kekayaan budaya Indonesia secara lebih mendalam serta menggambarkan nilai-nilai moral, tradisi, dan cara pandang yang khas. Setiap peribahasa memiliki asal-usulnya masing-masing dan sering kali terkait erat dengan kondisi sosial, budaya, atau alam di mana peribahasa tersebut berasal. Misalnya, peribahasa “Di mana



bumi dipijak, di situ langit dijunjung” mencerminkan nilai penghormatan terhadap lingkungan serta menunjukkan rasa hormat terhadap tempat asal atau lingkungan di sekitarnya.

Walaupun peribahasa merupakan bagian dari warisan budaya Indonesia, dalam era globalisasi dan digitalisasi, terdapat kesenjangan antara kemajuan teknologi dan pelestarian budaya. Hal ini mengakibatkan pemahaman terhadap peribahasa Indonesia semakin menurun di kalangan masyarakat, yang pada akhirnya dapat mengikis warisan budaya bangsa. Berdasarkan survei yang dilakukan oleh penulis, sebanyak 500 kuesioner telah dibagikan kepada 100 pelajar SMA, 100 pelajar SMP, 100 ASN, 100 pekerja swasta, dan 100 masyarakat umum. Kuesioner tersebut berisi 20 soal tentang peribahasa Indonesia. Berdasarkan hasil yang diperoleh, sebanyak 311 responden memiliki pemahaman peribahasa dengan nilai  $\leq 30$ , dan 189 responden memiliki nilai  $\geq 60$ . Survei ini membuktikan bahwa pemahaman masyarakat tentang peribahasa Indonesia masih sangat minim.

Berdasarkan permasalahan tersebut, untuk meminimalisir terkikisnya warisan budaya berupa pemahaman peribahasa, penulis menawarkan solusi dengan pendekatan Pencocokan Pola atau *Pattern Matching*. Pencocokan pola (*Pattern Matching*) dalam penelitian ini diimplementasikan menggunakan algoritma *Jaccard Similarity*. Algoritma ini berfungsi untuk mengukur kemiripan antara kumpulan kata kunci input dengan kumpulan kata kunci dari peribahasa-peribahasa dalam basis data dan sebagai metrik kemiripan leksikal karena kemampuannya dalam membandingkan kesamaan antara dua set data dengan menghitung rasio elemen-elemen yang sama terhadap total elemen unik. Oleh karena itu, pengembangan sistem interpretasi peribahasa Indonesia menggunakan *Pattern Matching* berbasis kata kunci merupakan sebuah tantangan dalam pelestarian budaya sekaligus wujud pemanfaatan teknologi untuk memahami kearifan lokal secara lebih luas dan berkelanjutan.

## 2. Kajian Literatur

### 2.1 Sistem Pakar

Sistem pakar adalah program komputer yang mensimulasikan proses berpikir seorang pakar untuk memecahkan masalah keputusan yang kompleks dalam domain tertentu. Berbagai definisi sistem pakar telah diajukan oleh beberapa penulis. Definisi umum yang mewakili fungsi yang dimaksudkan dari sistem pakar adalah: Sistem pakar adalah alat bantu pengambilan keputusan berbasis komputer interaktif yang menggunakan fakta dan heuristik untuk memecahkan masalah pengambilan keputusan yang sulit berdasarkan pengetahuan yang diperoleh dari seorang pakar. Sebagian besar aplikasi sistem pakar akan termasuk dalam salah satu dari kategori berikut: Interpretasi dan Identifikasi, Prediksi, Diagnosis, Perancangan, Perencanaan, Pemantauan, Debugging dan Pengujian, Instruksi dan Pelatihan, dan terakhir adalah Pengendalian.

Konsep sistem pakar untuk interpretasi peribahasa adalah membangun sebuah sistem yang meniru cara kerja seorang ahli dalam bidang bahasa. Sistem ini akan menangkap pengetahuan seorang pakar (seperti guru bahasa, budayawan, atau ahli sastra) tentang makna, konteks penggunaan, dan nilai-nilai budaya dalam peribahasa Indonesia, lalu menyimpannya dalam basis pengetahuan. Ketika pengguna memasukkan sebuah peribahasa atau kata kunci, sistem akan memprosesnya menggunakan mesin inferensi untuk memberikan interpretasi atau penjelasan yang mendekati penjelasan yang diberikan oleh seorang ahli.



## 2.2 Peribahasa

Lakoff dan Johnson dalam Nurgiyantoro (2007:83) mengemukakan bahwa manusia banyak menggunakan gaya bahasa dalam kehidupan sehari-hari untuk mengekspresikan berbagai keperluan. Peribahasa merupakan salah satu gaya bahasa yang berupa ungkapan tradisional dan merupakan salah satu wujud kebudayaan masyarakat. Dalam peribahasa terdapat unsur sistem budaya masyarakat yang berhubungan dengan nilai-nilai, pandangan hidup, norma, dan aturan-aturan dalam masyarakat.

Menurut Risma Devi, B. R. (n.d.) peribahasa dalam kutipannya dari (Foley, 1997: 361), Peribahasa tercermin suatu kearifan lokal atau *Folksy Wishdom* yang mengandung suatu moral atau kebenaran yang menjadi nilai-nilai yang dianut oleh suatu masyarakat. Sedangkan menurut Soebok (1962) melalui Foley bahwasanya peribahasanya bentuknya lebih cenderung menampakan suatu budaya dari suatu bangsa karena biasanya peribahasa diturunkan dari satu generasi ke generasi lain.

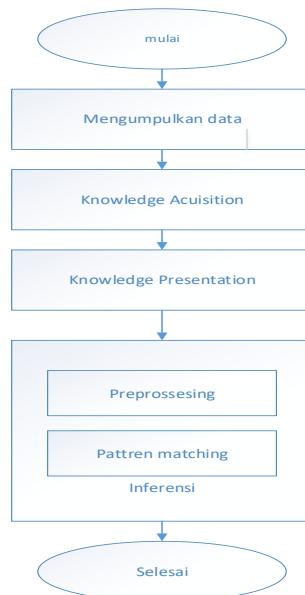
## 2.3 Patten Matching

*Pattern Matching* merupakan teknik pencocokan pola yang menggabungkan *keyword matching* dan *case-based logic*, dan dasarnya sangat mirip dengan seleksi kondisi tradisional. Secara definisi *Pattern Matching* merupakan metode yang digunakan untuk mencari dan mengidentifikasi keberadaan urutan karakter, kata, atau pola dalam data yang lebih besar. Metode ini sangat berharga dalam berbagai aplikasi seperti pengolahan teks, analisis data, dan pengembangan perangkat lunak, karena memfasilitasi pencarian dan manipulasi informasi secara efisien. Definisi lain dari *Pattern Matching* adalah teknik untuk mencocokkan pola tertentu dalam data dengan pola yang telah ditentukan sebelumnya. *Pattern Matching* dapat digunakan dalam berbagai bidang seperti pemrosesan teks, pencarian data, dan analisis data.

Penelitian ini *Pattern Matching* diimplementasikan menggunakan algoritma *Jaccard Similarity* sebagai core *matching engine* untuk mengukur tingkat kemiripan antara kata kunci input dengan *pattern* yang ada dalam *knowledge base*. *Jaccard Similarity* adalah metrik untuk mengukur kemiripan antara dua set data dengan membandingkan elemen-elemen yang sama terhadap total elemen unik dari kedua set tersebut

## 3. Metode Penelitian

Berikut metodologi penelitian dari interpretasi peribahasa Indonesia, diperlihatkan seperti pada gambar 1 di bawah ini:



**Gambar 1. Metodologi Penelitian Interpretasi**

### 3.1. Knowledge Acquisition

Sumber pengetahuan dalam penelitian ini diperoleh dari sumber-sumber tepercaya seperti: buku kumpulan peribahasa Indonesia, jurnal ilmiah kebahasaan, dan jurnal ilmiah sistem pakar. Data yang dikumpulkan berupa daftar peribahasa dan kata kunci utama yang mewakili setiap peribahasa sebagai *Key Pattern*. Proses knowledge acquisition meliputi:

- Pengumpulan Data: Mengumpulkan peribahasa Indonesia dari sumber tepercaya
- Identifikasi Kata Kunci: Menentukan kata kunci unik yang mewakili setiap peribahasa, dengan fokus pada kata benda, kata kerja atau kata sifat
- Formulasi Interpretasi: Menuliskan makna, nasihat, atau pesan moral dari setiap peribahasa dengan bahasa yang sederhana dan mudah dipahami

### 3.2. Knowledge Representation

Pengetahuan yang ada direpresentasikan dalam bentuk *array of object* dalam kode program. Bentuk representasi berupa tabel, seperti contoh pada tabel 1 di bawah, terdiri dari Id, Peribahasa Indonesia, Kata Kunci di setiap masing-masing peribahasa Indonesia dan Interpretasi setiap peribahasa.

**Tabel 1. Knowledge Representation**

ID	Peribahasa	Kata Kunci Patten	Interpretasi
P01	Air Tenang Menghanyutkan	[“Air”, “Tenang”, “Hanyut”]	Orang yang kelihatannya pendiam, tetapi memiliki ilmu atau kemampuan yang hebat.
P02	Bagai Air di Daun Talas	[“Air”, “Daun”, “Talas”]	Seseorang yang tidak punya pendirian tetap atau plin-plan.
P03	Besar Pasak Daripada	[“Besar”, “Pasak”, “Tiang”]	Pengeluaran lebih



	Tiang		besar daripada pendapatan
P04	Tak Ada Gading Yang Tak Retak	[“Gading”, “Retak”]	Tidak ada sesuatu yang sempurna di dunia
...	...	...	...

### 3.3 Inferensi dengan Pattern Matching Berbasis Kata Kunci

Tahap inferensi dilakukan untuk menarik kesimpulan berdasarkan kata kunci yang diinputkan pengguna. Proses ini terdiri dari dua tahap utama:

#### a. Preprocessing Input Teks

Tahap preprocessing dilakukan untuk mempersiapkan input teks sebelum proses matching:

1. *Case Folding*: Mengubah semua karakter input menjadi huruf kecil
2. *Tokenization*: Memecah kalimat input menjadi kata-kata individual
3. *Filtering*: Menghilangkan kata-kata umum yang tidak penting (*stop words*)

#### b. Pattern Matching dan Scoring

Tahap pattern matching dan scoring ini membandingkan kata kunci input dengan *pattern* di basis pengetahuan:

1. Pattern Matching dengan Jaccard Similarity:

Menggunakan algoritma Jaccard Similarity untuk mengukur kemiripan antara dua set data.

$$\text{Rumus: } J(A,B) = |A \cap B| / |A \cup B| \quad (1)$$

Dimana:

- $A \cap B$  = Irisan (*intersection*) - elemen yang ada di A dan B
- $A \cup B$  = Gabungan (*union*) - semua elemen unik dari A dan B
- $J(A,B)$  = Nilai similarity antara 0 dan 1

2. Scoring System:

- Memberikan skor untuk setiap peribahasa berdasarkan jumlah kata kunci yang cocok
- Skor akhir merupakan kombinasi antara Jaccard Similarity dan persentase kata kunci yang match

### 3.4. Evaluasi Sistem

Evaluasi dilakukan untuk mengukur kinerja sistem dengan metrik:

1. *Precision*: Persentase peribahasa yang relevan dari hasil yang dikembalikan
2. *Recall*: Kemampuan sistem menemukan semua peribahasa yang relevan
3. *F1-Score*: Harmonic mean antara precision dan recall

## 4. Hasil dan Pembahasan

### 4.1 Knowledge Base

Basis data dari peribahasa Indonesia terdiri dari peribahasa Indonesia populer beserta kata kunci dan interpretasinya:

**Tabel 2. Knowledge Base**

No	Peribahasa	Jumlah Kata Kunci	Kategori
1	Air tenang menghanyutkan	3 kata	Sikap dan Perilaku
2	Bagai air di daun talas	3 kata	Sikap dan Perilaku
3	Besar pasak daripada tiang	3 kata	Ekonomi



4	Tak ada gading yang tak retak	2 kata	Filsafat
..	...	...	...

#### 4.2 Hasil Pengujian

Pengujian dilakukan dengan menginputan kata kunci dengan empat cara. Inputan dengan kata kunci lengkap, kata kunci parsial, kata kunci sinonim, kata kunci negative (kata kunci yang tidak relevan). Dari empat kategori kata kunci yang diinputkan, berdasarkan proses pencocokan kata kunci dengan algoritma *jaccard similarity* menggunakan persamaan 1, hasilnya: inputan kata kunci lengkap, scor kecocokan > 0.80, inputan parsial, scor kecocokan > 0.60, inputan kata kunci sinonim, scor kecocokan > 0.40 dan inputan kata kunci negative, scor kecocokan adalah 100%

```
-----  
Masukkan teks yang ingin diinterpretasi: air di daun talas  
=====  
HASIL INTERPRETASI PERIBAHASA  
=====  
Input teks: air di daun talas  
-----  
PERIBAHASA TERKAIT: Bagai air di daun talas  
SKOR KECOCOKAN: 50.00%  
INTERPRETASI: Seseorang yang tidak punya pendirian tetap atau plin-plan.  
KATA KUNCI YANG COCOK: daun, talas, air  
  
ALTERNATIF LAIN:  
1. Air tenang menghanyutkan (Skor: 12.50%)  
2. Sambil menyelam minum air (Skor: 12.50%)  
=====
```

**Gambar 2. Hasil Pengujian**

Kata kunci: air di daun talas, merupakan kata kunci parsial.

#### Tahap pertama: Proses preprocessing

# Tokenization & Stop Words Removal

Input asli: "air di daun talas"

Setelah preprocessing: ["air", "daun", "talas"]

Stop words removed: ["di"]

#### Tahap kedua perhitungan *Jaccard Similarity* :

##### Langkah ke-1: Hitung *Intersection* (Kata yang Cocok)

input\_tokens = {"air", "daun", "talas"}

peribahasa\_keywords = {"air", "daun", "talas", "bolak", "balik", "tidak tetap"}

intersection = input\_tokens ∩ peribahasa\_keywords

= {"air", "daun", "talas"}

|intersection| = 3

##### Langkah ke-2: Hitung Union (Semua Kata Unik)

union = input\_tokens ∪ peribahasa\_keywords



= {"air", "daun", "talas", "bolak", "balik", "tidak tetap"}  
|union| = 6

### Langkah ke-3: Hitung Jaccard Similarity

Jaccard Similarity = |intersection| / |union|  
= 3 / 6

untuk input “air di daun talas” yang menghasilkan skor **50.00%**:

#### 4.3 Analisa akurasi

Untuk menguji akurasi interpretasi dilakukan pengujian dengan 30 kalimat uji yang mewakili berbagai skenario:

Kategori	Jumlah Test Case	Akurasi
Input dengan kata kunci eksak	10	90%
Input dengan kata kunci parsial	10	70%
Input dengan sinonim	5	40%
Input negative (kata kunci yang tidak relevan)	5	100%

Rata-rata akurasi nya adalah 76% diperoleh dari perhitungan:  $((10*90) + (10*70) + (5*40) + (5*100)) / 30 = (900+700+200+500)/30 = 2300/30 = 76.67\%$  berdasarkan dari empat kategori kata kunci yang diinputkan Untuk input negatif nilai akurasi 100% berarti sistem selalu benar dalam tidak merekomendasikan apa-apa.

## 5. Kesimpulan

Kesimpulan dari penelitian ini adalah algoritma *Jaccard Similarity* mampu untuk mencocokan kata kunci dari setiap peribahasa Indonesia di database dengan tiga kategori kata kunci. Hasil uji coba rata-rata akurasi kecocokan 70% dari 30 skenario yang dicoba. Mampu menangkap kemiripan konsep meski dengan kata yang berbeda, cukup fleksibel untuk variasi input pengguna, komputasi efisien untuk aplikasi real-time. Dan proses *preprocessing* mampu berhasil membersihkan noise. Keterbatasan utama model saat ini adalah akurasi rendah 40% dalam menangani input kata kunci sinonim, mengindikasikan bahwa penelitian selanjutnya perlu mengintegrasikan metode pencocokan berbasis semantic. Penelitian ini membuka peluang untuk pengembangan lebih lanjut dalam bidang *natural language processing* untuk bahasa Indonesia dan preservasi budaya.

## Daftar Pustaka

- Astuti Winarti, A. R. (2023). Peribahasa Favorit menurut Perspektif Pemaknaan Objektif Siswa Kelas XI SMANegeri 3 Sintang. *Jurnal Pembelajaran Bahasa dan Sastra*, 537-544
- Haeruddina, S. C. (2025). Evaluation of Chatbot Responsiveness and Accuracy Using a Matching Pattern Algorithm for Implementation in a Medium-Sized Enterprise. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 126-132.
- Helena Nurramdhani Irmanda, R. A. (2021). Pembuat Sampiran Pantun Otomatisberbasis Pattern-matching. *Jurnal SISFOKOM(Sistem Informasi dan Komputer)*, 306-311.
- Juniati, S. (2021). ANALISIS STRUKTUR FUNGSIONAL PADA PERIBAHASA INDONESIA. *CENDEKIA: Jurnal Ilmiah Pendidikan*.



- Liga Mayola, M. D. (2024). Algoritma jaccard Similarity Untuk Deteksi Keiripan Judul Desertasi Dengan Pendekatan Variasi Stop Word Removal. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 477- 478.
- Migunani, K. A. (2020). PEMANFAATAN NATURAL LANGUAGE PROCESSING DAN PATTERN MATCHING DALAM PEMBELAJARAN MELALUI GURU VIRTUAL. *Jurnal elektronika dan Komputer*, 121 - 133.
- Mohamad Ilyas Abas, D. W. (2023). Tinjauan Literatur Sistematis Sistem Pakar. *Terapan Informatika Nusanta*, 497-502.
- MOSINGALETE LUMBU JUDITH, M. N. (2020). A Machine Learning for an Expert System on Behalf of Decision Making: Application to Lumbago. *International Journal of Computer Science Issues (IJCSI)*, 27-33.
- Muhammad Rizal Waskito, A. D. (2024). IMPLEMENTASI METODE COLLABORATIVE FILTERING MENGGUNAKAN ALGORITMA COSINE SIMILARITY DAN JACCARD SIMILARITY PADA SISTEM E-COMMERCE. *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, 4307-4316.
- Nugraheni, M. (2020). PERBANDINGAN JACCARD SIMILARITY DENGAN EXTENDED. *Jurnal Pinter*.
- Paness, I. Y. (2021). SITEKTUR SISTEM PAKAR: PENGENALAN SISTEM PAKA. PT. LAMINTAN.
- Roni Putra, S. Z. (2024). Similarity Judul Tugas Akhir Menggunakan Metoda Cosine, Jaccard, Rabin-Karp Pada Jurusan TI. *Jurnal Ilmiah Teknologi Sistem Informasi (Jitsi)*, 206-212.
- Sarmiana Naomi, R. R. (2025). Systematic Literature Review dengan PRISMA: Peran Peribahasa Indonesia dalam Pembentukan Karakter Generasi Z di Era Digital. *LENTERA JURNAL PENGABDIAN MASYARAKAT*.
- Sirait, H. (2023). Basic Elements and Characteristics in Building an Expert System. *Journal Neosantara Hybrid Learning*, 227-248.
- Susi Rianti, R. A. (2023). PERBANDINGAN ALGORITMA EDIT DISTANCE, LEVENSSTEIN DISTANCE, HAMMING DISTANCE, JACCARD SIMILARITY DALAM MENDETEKSI STRING MATCHING. *JSI (Jurnal Sistem Informasi)*, 305.
- Vina Larasati, ., P. (2025). Chatbot Model Design Based On Pattern Matching For Information Services About Pest Organisms In Rice Fields . *Journal of Artificial Intelligence and Software Engineering*, 87-95.