

OPTIMASI MESIN PENCARI BUKU FIKSI BERDASARKAN PADA SEMANTIK IMPRESI

Varidh Bimantara¹, Rengga Asmara², Nur Rasyid Muhtada³

^{1,2,3}Departemen Teknik Informatika dan Komputer,
Politeknik Elektronika Negeri Surabaya
, Jl. Raya ITS, Kampus ITS, Keputih, Sukolilo, Surabaya

ABSTRACT

Fiction books are one of the most popular books in Indonesia. There are five most popular genres in fiction books, namely fantasy, mystery, romance, sci-fi and thriller. Each genre gives a different impression and its own fans for the reader. It is common practice when people choose fiction books based on the title, author or publisher of the book. However, this does not provide precise search results. In this final project, an application system was developed to find out fiction books based on the semantic impressions contained on the cover of the fiction book. The impression on each book cover is obtained through a survey of fiction book lovers in Indonesia. To get the results of the proximity between the user search and the impression survey data obtained through text mining, as well as the cosine similarity algorithm to calculate the most precise proximity value to the impression expected by the user. The results of this system display fiction books that have the most precise proximity value to the impression expected by the user.

Keywords: *Fiction, impression, mining text, cosine similarity*

I. INTRODUCTION

Saat ini ketika seseorang melakukan pencarian buku fiksi masih mengalami kesulitan dikarenakan kurangnya presisi dari hasil buku fiksi yang di cari. Sejauh saat ini kebanyakan orang mencari buku fiksi berdasarkan judul, pengarang, maupun penerbit. Namun hal tersebut terkadang masih belum mewakili cerita yang diinginkan oleh pembaca. Ada permasalahan jika seorang pembaca merupakan orang yang baru ingin membeli buku fiksi, belum memiliki pengalaman dalam memilih atau belum banyak mengetahui tentang judul atau pengarang buku fiksi mana yang biasanya memiliki cerita menarik untuk dibaca. Dilain sisi sampul buku fiksi memberikan kesan terhadap isi dari keseluruhan buku fiksi tersebut. Sampul pada buku dapat mewakili synopsis cerita yang dituangkan dalam bentuk gambar maupun typografi pada sampul buku fiksi. Sampul buku merupakan hal pertama yang akan dilihat seorang pembaca, begitu juga dengan buku fiksi. "Don't judge the book by its cover" nampaknya bukan lagi menjadi idiom yang tepat untuk menggambarkan proses seorang pembaca memilih buku yang akan ia baca. Nyatanya, sampul buku punya peran penting dalam proses penerbitan sebuah buku, yang di dalamnya meliputi penempatan obyek ilustrasi yang tepat, pemilihan font yang nyaman dibaca, hingga penggunaan warna yang sesuai dengan emosi yang ingin penulis sampaikan lewat bukunya, dan yang penting adalah koherensi dengan cerita. Sebagai sebuah produk komoditas, tentu saja buku dengan pewajahan yang menarik meningkatkan daya jualnya.

Melalui jurnal dari Tess Adams " *Judging a Book By Its*

Cover: Are First Impressions Accurate?" mengatakan bahwasanya ada korelasi yang timbul dari sebuah sampul buku dengan impresi. Melalui sampul buku, bisa didapatkan sebuah informasi tersirat^[2]. Selain itu melalui jurnal Assist. Prof. Dr. Asli Sezgin "The First Impression of "Best Selling" Agenda Books: Semiotic Analysis of Book Covers" menyampaikan dari hasil penelitian yang dilakukannya menyebutkan bahwasanya pada setiap sampul buku memiliki impresi masing-masing^[3]. Hal ini bisa menjadi sebuah rekomendasi cara baru bagi seorang pembeli buku dalam mencari buku yang diinginkan.

II. RELATED WORKS

Aplikasi Pencarian Buku Berbasis Web Semantik Untuk Perpustakaan Smk Yadika 7 Bogor^[4], Penelitian yang dilakukan oleh Javier Zebua dan Metty Mustikasari mengatakan bahwa pada aplikasi pencarian berbasis web semantik ini ditujukan untuk mempermudah pengunjung perpustakaan di SMK Yadika 7 Bogor dalam melakukan pencarian buku. Penelitian ini dilakukan dengan mengumpulkan data dan informasi dari perpustakaan serta merangkum teori mengenai web semantik dan unsur-unsur yang dibutuhkan dalam pembuatan aplikasi ini. Pembuatan aplikasi ini dimulai dengan perancangan ontologi yang digunakan dalam pembuatan struktur semantik perpustakaan. Struktur ontologi buku pada penelitian ini meliputi judul, penulis, penerbit, kategori, tahun terbit, jumlah, klasifikasi, asal buku, nomor induk, nomor rak dan keterangan. Struktur tersebut selanjutnya diubah dalam bentuk RDF yang merupakan salah satu struktur bahasa ontologi. Selanjutnya dilakukan perancangan *query* dengan menggunakan *query* SPARQL untuk mengambil informasi dari RDF. Dilanjutkan dengan perancangan tampilan aplikasi. Setelah tahap perancangan, dilanjutkan dengan tahap implementasi yaitu implementasi *query* dan implementasi *interface*. Setelah

aplikasi selesai dikerjakan, akan dilakukan ujicoba dan aplikasi akan dievaluasi oleh pengguna.

Aplikasi Pencarian Buku Berbasis Web Semantik Untuk Perpustakaan Universitas Muhammadiyah Yogyakarta [5], Penelitian yang dilakukan oleh Siti Khotcitah mengatakan pada penelitian ini bahwa semantik Web adalah seperangkat teknologi, yang memungkinkan komputer untuk memahami makna informasi berdasarkan metadata, yaitu informasi tentang informasi. *Semantic-Web-Based Searching Application For Doctors Schedule And Facilities In Hospital*[6], Penelitian yang dilakukan oleh A.B. Mutiara mengatakan pada penelitian ini bahwa Rumah Sakit sebagai media pendukung kesehatan memiliki informasi utama seperti jadwal praksis dokter yang ada dan ketersediaan fasilitas. Teknologi yang digunakan adalah mesin pencari berbasis web semantik ontologis. Metode web semantik dengan pendekatan ontologi tidak hanya mampu memahami makna kata dan konsep, tetapi juga hubungan logis di antara mereka. Dalam tulisan ini akan dijelaskan pengembangan mesin pencari dengan dua jenis data, yang diambil langsung (live data) dengan menggunakan konsep Ontology Web Language (OWL) dan secara manual dimasukkan (data dummy) dengan menggunakan konsep Resource Kerangka Deskripsi (RDF).

Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi[7], Penelitian yang dilakukan oleh Rizki Tri Wahyuni mengatakan pada penelitian ini bahwa Metode TFIDF merupakan suatu cara untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Metode cosine similarity merupakan metode untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah vector dengan menggunakan keywords (kata kunci) dari sebuah dokumen sebagai ukuran. Metode pengembangan sistem yang digunakan dalam penelitian ini adalah model waterfall, sedangkan metode penelitian yang digunakan adalah metode Research and Development (R&D). Data latih yang digunakan dalam penelitian ini berjumlah 50 dokumen skripsi dengan beberapa kategori yang berbeda. Hasil penelitian menunjukkan bahwa persentase tingkat ketepatan klasifikasi sistem adalah sebesar 98%.

III. SYSTEM DESIGN

Untuk membangun system pencarian buku fiksi bedasarkan semantic impresi ini penulis menggunakan alur proses seperti yang ditunjukkan pada figure 1.

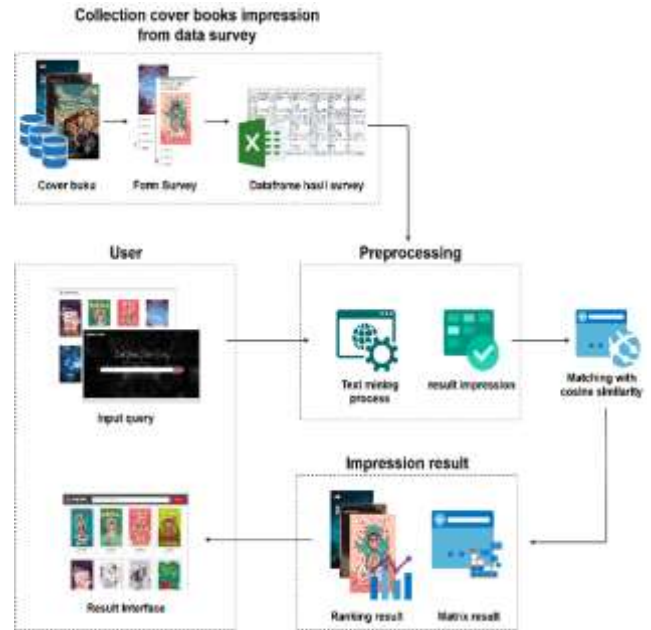


Figure 1. System Design

Pengumpulan Impresi Sampul Buku dari Survey

Pada proses ini bermula dari data yang telah dikumpulkan berupa beberapa cover buku fiksi. Dari cover buku fiksi tersebut akan dibuat form survey, dimana isi dari form survey tersebut adalah cover buku fiksi dan pilihan multiple choice impresi. Form survey ini nantinya akan di sebar kepada sepuluh komunitas pecinta buku yang ada pada facebook.

Preprocessing Query

Pada tahap ini, query yang di inputkan oleh user akan diproses pada teks mining. Teks mining disini akan memproses query yang di inputkan oleh user dan dataframe impresi yang diperoleh dari survey. Dalam proses teks mining ada beberapa tahapan yang dilakukan yakni :

a. Tokenizing

Pada tahap tokenizing, dataframe impresi dan query impresi dari user akan dijadikan satu dataframe. Dataframe pada impresi survey hanya akan diambil impresinya saja. Query impresi yang di inputkan oleh user akan masuk kedalam satu dataframe dengan letak baris paling bawah. Semua kata-kata akan dibaca perbaris, kata-kata tersebut akan dipisah dengan karater spasi sebagai pemisahannya. Sehingga hasil yang keluar dari proses filtering ini adalah setiap kata akan terpisah pada setiap baris.

b. Filtering

Pada tahap filtering disini merupakan tahap dimana mengambil kata-kata penting dengan menggunakan algoritma stopwords. Pada setiap baris hasil output dari tokenizing akan dicek setiap kata, jika kata tersebut terdapat pada dataframe stopwords maka kata tersebut akan dihapus, dan jika kata tersebut

tidak ada pada dataframe stopwords maka kata tersebut akan disimpan pada table koleksi.

c. Stemming

Pada tahap stemming disini merupakan tahap untuk mendapatkan kata dasar dari setiap kata hasil filtering. Proses dilakukan dengan menganalisa setiap baris, jika kata pada setiap baris tersebut mengandung awalan atau akhiran maka kata awalan atau akhiran tersebut akan di hapus, sehingga output dari proses ini adalah kata dasar dari setiap kata pada masing-masing baris.

d. Pembobotan Tf-Idf

Pada tahap pembobotan Tf-Idf disini, semua kata output dari proses teks mining akan dihitung banyaknya pada setiap baris. Baris disini dimaksudkan setiap kepemilikan sampul buku. Setiap baris akan dihitung nilai impresi yang diperoleh oleh masing-masing sampul buku. Hasil pembobotan dari table Tf diatas selanjutnya akan dikalikan dengan nilai Idf dengan rumus untuk nilai Idf :

$$IDF_j = \log\left(\frac{D}{df_j}\right)$$

Matching with Cosine Similarity

Pada tahap ini, matriks dari impresi data survey akan dilakukan matching dengan kata impresi query yang di inputkan oleh user. Proses matching disini menggunakan metode cosine similarity. Jadi dari query akan di hitung nilai kedekatannya dengan nilai yang ada pada matriks impresi dengan menggunakan rumus :

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

- A = vector dokumen
- B = vector query
- A . B = perkalian dot vector A dan vector B
- |A| = panjang vector A
- |B| = panjang vector B
- |A| |B| = cross product antara |A| dan |B|

Dari proses pencarian nilai cosine similarity diatas maka didapatkan masing-masing nilai cosine similarity dari tiap sampul buku.

Impression Result

Hasil dari proses matcing dengan menggunakan metode cosine similarity adalah matriks dua dimensi, dimana pada

masing-masing fitur menyimpan nilai kedekatan antara query dengan impresi pada cover buku. Hasil matriks tersebut akan dilakukan sorting berdasarkan nilai tertinggi, dan kemudian ditampilkan pada layar user. Tampilan output yang dikirim pada user adalah berupa list cover buku fiksi yang memiliki kedekatan dengan query yang di inputkan oleh user.

IV. EXPERIMENT AND ANALYSIS

Penulis akan menguji ketepatan hasil pencarian dengan data hasil survey. Dilakukan perbandingan hasil pencarian menggunakan kedekatan cosine similarity dengan impresi dengan nilai tertinggi sesuai dengan impresi query yang di inputkan oleh user. Data impresi diperoleh dari survey yang dilakukan kepada komunitas pecinta buku fiksi yang ada di Indonesia. Hasil data survey menunjukkan nilai jumlah impresi dari sampel 100 buku yang telah disebar untuk mendapatkan penilaian impresi. Dari hasil survey menunjukkan hasil impresi pada survey tersebut sesuai dengan impresi yang terkandung pada genre buku yang bersangkutan. Dengan demikian data tersebut telah sesuai dengan impresi yang ada pada genre buku itu sendiri.

Setelah mendapatkan nilai Tf-Idf masing-masing impresi pada setiap sampul buku, maka dari nilai tersebut akan masuk pada perhitungan cosine similarity untuk mendapatkan nilai paling besar atau berarti paling dekat dengan query impresi yang di inputkan oleh user. Terdapat 33 impresi yang didapatkan dari survey. Dari 33 percobaan disetiap impresi mendapatkan hasil seperti tabel berikut :

Tabel 1. Error rate

No	Impresi	Error Rate
1	Aneh	0%
2	Berani	0%
3	Canggih	50%
4	Cemas	0%
5	Estetik	0%
6	Fantastis	0%
7	Futuristis	0%
8	Historis	0%
9	Ilmiah	0%
10	Imajinatif	0%
11	Irasional	0%
12	Kasar	0%
13	Kejam	0%
14	Khayal	0%
15	Kontemporer	0%
16	Lampau	0%
17	Logis	0%
18	Lucu	0%
19	Magis	0%
20	Misterius	0%
21	Modern	0%
22	Ngeri	0%
23	Patriotik	0%
24	Penasaran	0%
25	Politis	0%
26	Psikologis	0%
27	Romantis	0%

28	Sedih	80%
29	Spekulatif	0%
30	Supernatural	0%
31	Takut	0%
32	Tegang	0%
33	Tradisional	0%

- [7] Rizki Tri Wahyuni, “Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi”, Universitas Negeri Semarang, 2017.

Berdasarkan pada tabel 4.40. diatas dapat disimpulkan untuk nilai error rate dari keseluruhan percobaan pada masing – masing impresi adalah sebesar 3,93%.

V. CONCLUSIONS

Pencarian buku fiksi Indonesia dapat dilakukan dengan berbagai macam cara, termasuk salah satunya dengan menggunakan metode semantik impresi. Semantic impresi adalah metode untuk mencari buku fiksi berdasarkan impresi yang terkandung pada cover buku tersebut.

Impresi pada masing-masing buku fiksi didapatkan melalui survey kepada komunitas pecinta buku di Indonesia. Hasil dari data survey tersebut akan melalui beberapa tahap untk pada akhirnya mendapatkan hasil yang diinginkan. Tahap pertama dari semantic impresi ini adalah dengan menggunakan metode text mining. Pada tahap text mining disini berawal dari tokenizing, filtering, stemming dan pembobotan Tf-Idf. Beberapa tahap yang ada di metode text mining tersebut akan dilanjutkan pada pencarian penfekatan menggunakan metode cosine similarity. Setiap buku yang diajukan pada data survey, disini ada 100 buku akan memiliki nilai cosine similarity ata nilai kedekatan masing-masing terhadap query impresi yang di inputkan oleh user. Selanjutnya nilai tersebut akan di sorting berdasarkan nilai terbesar untuk mendapatkan hasil yang paling dekat dengan queri impresi.

Proses semantic impresi buku fiksi Indonesia menghasilkan nilai error rate sebesar 3,93 %. Hasil tersebut didapatkan dari percobaan terhadap keseluruhan impresi yang ada.

REFERENCES

- [1] Aditya Jaya Iswara, “Jenis Buku Apa yang Paling Laris di Indonesia” 8 Juni 2019.
- [2] Tess Adam, “Judging a Book By Its Cover: Are First Impressions Accurate?” Spring 2012, Boulder AS, University of Colorado Boulder, 2012.
- [3] Assist. Prof. Dr. Asli Sezgin, “The First Impression of "Best Selling" Agenda Books: Semiotic Analysis of Book Covers”, Osmaniye Turkey, Osmaniye Korkut Ata University, 2014.
- [4] Javier Zebua, “Aplikasi Pencarian Buku Berbasis Web Semantik Untuk Perpustakaan Smk Yadika 7 Bogor”, Universitas Gunadarma, 2012.
- [5] Siti Khoctitah, “Aplikasi Pencarian Buku Berbasis Web Semantik Untuk Perpustakaan Universitas Muham madiyah Yogyakarta”, Universitas Muhammadiyah Yogyakarta, 2015.
- [6] A.B. Mutiara, “Semantic-Web-Based Searching Application For Doctors Schedule And Facilities In Hospital”, Universitas Gunadarma, 2014.