

EKSTRAKSI INFORMASI UNTUK NOVEL

Arbi Insani¹, Nelly Indriani Widiastuti²

^{1,2} Teknik Informatika – Universitas Komputer Indonesia

Jalan Dipatiukur 112-116 Bandung

E-mail : ¹ arbi_benai@yahoo.com, ² nelly.indriani@email.unikom.ac.id

ABSTRAK

Informasi dalam bentuk digital menjadi salah satu alternatif untuk mendapatkan informasi yang akurat dan cepat. Pada umumnya setiap perpustakaan melakukan pendataan informasi dan identitas suatu dokumen dilakukan dengan cara mengisi data kedalam sistem, masalah yang dihadapi oleh pustakawan adalah setiap kurun waktu ke waktu dokumen setiap tahunnya bertambah dan jumlahnya yang banyak mengakibatkan dalam melakukan pendataan informasi novel banyak terjadi kesalahan dan membutuhkan waktu yang lama dalam pendataan novel. Dalam penelitian ini dibangun suatu sistem agar dalam pengerjaan pendataan informasi dapat dilakukan secara otomatis oleh sistem yaitu dengan cara mengekstraksi informasi tersebut secara mudah, cepat dan menghasilkan data informasi yang benar. Ekstraksi informasi yang digunakan dalam penelitian ini menggunakan metode berbasis aturan atau disebut *rule-based*. Metode ini merupakan suatu metode yang menggunakan aturan berdasarkan fakta dari data yang dianalisis. Hasil penelitian ini menunjukkan hasil setelah dilakukan pengujian akurasi menggunakan 50 buku berupa novel. Hasil yang ditunjukkan sangat baik. Hal ini berarti ekstraksi informasi pada penelitian ini dapat digunakan untuk mengekstrak identitas data pada novel.

Kata Kunci : *Rule-based*, ekstraksi Informasi, novel.

1. PENDAHULUAN

Proses perkembangan teknologi informasi dan komunikasi saat ini begitu pesat dan meningkat, terutama informasi dalam bentuk novel, jurnal, majalah, artikel, koran, dan berita. Informasi dalam bentuk digital menjadi salah satu alternatif untuk mendapatkan informasi yang akurat dan cepat. Novel adalah karangan yang panjang dan berbentuk prosa dan mengandung rangkaian suatu cerita kehidupan seseorang dengan orang lain maupun

cerita fiksi yang ada di sekelilingnya dengan menonjolkan watak dan sifat setiap pelaku [1]. Perkembangan novel selalu bertambah setiap waktunya seperti contoh novel manusia setengah salmon, 7 rahasia shope dan lain sebagainya. Pada umumnya setiap perpustakaan melakukan pendataan informasi dan identitas suatu dokumen novel dilakukan dengan cara mengisi data kedalam sistem. Masalah yang dihadapi oleh pustakawan adalah dari waktu ke waktu dokumen setiap tahunnya bertambah. Jumlah yang banyak mengakibatkan terjadi banyak kesalahan dalam melakukan pendataan informasi novel dan membutuhkan waktu yang lama.

Proses ekstraksi informasi (*Information Extraction* atau IE) adalah sebuah proses untuk menemukan informasi terstruktur dari dokumen tidak terstruktur atau semi terstruktur. Adapun data yang terstruktur adalah data yang telah terorganisir sehingga mudah dalam suatu pencarian data sedangkan tidak terstruktur adalah data yang belum terorganisir [2]. Contoh dari ekstraksi informasi adalah mengekstraksi iklan Koran, berita, karya ilmiah dan lain sebagainya sesuai penelitian sebelumnya oleh Ramón yaitu menggunakan ekstraksi informasi untuk mengklasifikasikan iklan koran [3]. Penelitian yang dilakukan Riaddy dkk. membahas ekstraksi informasi untuk makalah ilmiah menggunakan pendekatan supervised learning [4]. Pada penelitian lain menyimpulkan bahwa ekstraksi informasi pada dokumen LHP (Laporan Hasil Pemeriksaan) dapat dilakukan dengan menggunakan *rule-based classification* dengan akurasi data latih sebesar 98,27% dan akurasi data uji sebesar 89,77% [5]. Penelitian oleh Muhammad Habibi yang melakukan ekstraksi informasi untuk evaluasi pemahaman fisika kinematika bisa dilakukan dengan metode *rule-based* dengan hasil akurasi sistem yang telah dibangun menghasilkan nilai F – Measure berkisar antar 90.6% - 95.4% [6]. Pada penelitian oleh M. Karibun menyimpulkan bahwa penerapan metode *rule-based* dengan *unsupervised learning* untuk pelabelan dokumen berbahasa Indonesia dapat mengenali 97,85% token [7]

Berdasarkan uraian tersebut maka dalam penelitian ini bertujuan untuk membangun sebuah

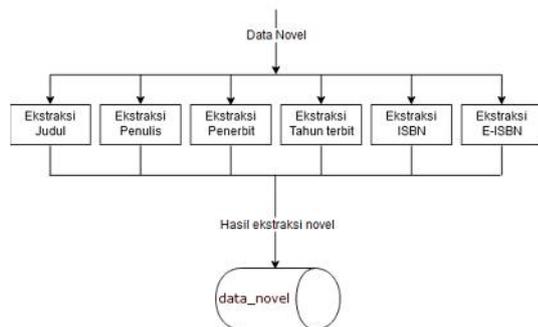
sistem ekstraksi informasi dokumen teks novel dengan menggunakan metode *rule-base* agar dapat dalam pendataan informasi menghasilkan pendataan yang akurat, cepat, mudah dan tidak membutuhkan waktu yang lama. Dengan adanya sistem tersebut diharapkan akan lebih mudah dan cepat dalam pendataan informasi novel.

2. ISI PENELITIAN

2.1 Metode Penelitian

Pada penelitian ini sistem dibangun menggunakan metode *rule-base* untuk ekstraksi informasi dokumen. Dimana *rule-base* dikenal juga dengan sebutan pendekatan *hand-crafting* atau Knowledge engineering. Disebut pendekatan rule base karena rule-base ini memanfaatkan keahlian manusia (*human expert*) untuk membuat suatu aturan-aturan (rules) secara manual melalui proses pemahaman manusia itu sendiri. *Human expert* atau pakar dituntut untuk bisa memahami domain sistem [8].

Adapun gambaran umum analisis sistem yang akan di bangun pada ekstraksi informasi dokumen teks ini dapat dilihat pada gambar 1.



Gambar 1. Gambaran Sistem

Gambaran Tahapan Sistem akan dijelaskan sebagai berikut ini.

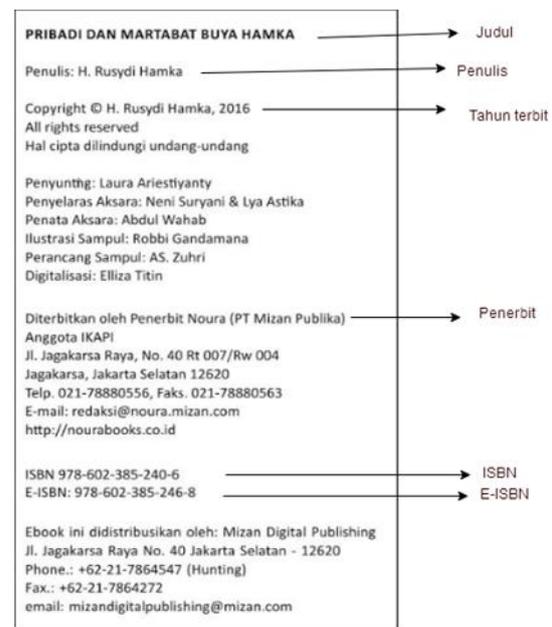
1. Pada tahap awal pengguna memasukan file teks yang merupakan novel yang berformat .txt ke dalam sistem
2. Selanjutnya sistem akan mengekstraksi informasi yaitu judul, penulis, penerbit, tahun terbit, ISBN, E-ISBN dari data novel tersebut. Proses ini dilakukan dengan mencari kata kunci yang dapat dijadikan acuan pembentukan aturan.
3. Tahap terakhir informasi dari data akan disimpan ke dalam tabel.

2.2 Analisi Data Masukan

Analisis data masukan pada penelitian ini menggunakan data novel yang berformat jpg yang di konversi ke .txt. menggunakan aplikasi OCR atau *Optical Character Recognition*. OCR adalah aplikasi yang berfungsi untuk men-scan gambar kemudian dikonversi menjadikan text dan aplikasi ini juga dapat menjadi *support* /aplikasi tambahan untuk

scanner. Dengan adanya OCR, Image yang bertulisan tangan, tulisan mesin ketik atau computer text, dapat dimanipulasi kedalam bentuk format yang berbeda [9].

Pada penelitian ini analisis berfokus pada pencarian kata kunci dan aturan yang akan digunakan untuk mengekstraksi informasi data novel. Berikut adalah data novel sebelum di konversi ke .txt dilihat pada gambar 2 berikut.



Gambar 2. data novel

Pada gambar 2 adalah contoh data novel yang berjudul “PRIBADI DAN MARTABAT BUYA HAMKA” dan akan diambil 6 data ekstraksi informasi yaitu :

1. Judul novel yaitu “PRIBADI DAN MARTABAT BUYA HAMKA” yang muncul pada awal data novel sehingga hal ini dapat menjadi aturan dalam pengambilan ekstraksi informasi.
2. Penulis yaitu “H. Rusydi Hamka”. Pengambilan penulis di peroleh dari kata kunci dikarenakan sebelum kata H. Rusydi Hamka terdapat kata kunci yaitu penulis
3. Tahun terbit yaitu “2016” merupakan tahun terbit dari novel. Pengambilan tahun terbit menggunakan rentang nilai angka 2001 – 2025.
4. Penerbit yaitu “Noura (PT Mizan Publika)” merupakan penerbit dari novel untuk pengambilan penerbit menggunakan kata kunci dikarenakan sebelum kata noura terdapat kata penerbit.
5. Nomor ISBN yaitu “973-602-385-240-6” merupakan ISBN novel. Pengambilan ekstraksi menggunakan kata kunci karena sebelum 973-602-385-240-6 terdapat kata ISBN.

6. Nomor e-ISBN yaitu “978-602-385-246-8” merupakan kode e-ISBN novel. Pengambilan ekstraksi menggunakan kata kunci karena sebelum 978-602-385-246-8 terdapat kata e-ISBN

2.3 Analisis Pembentukan Aturan dan Kata Kunci

Kunci

Kata kunci adalah suatu kata yang menjadi acuan suatu dari data [10]. Kata kunci dapat digunakan sebagai acuan dalam pembentukan aturan untuk pengambilan suatu data dalam proses ekstraksi informasi pada novel. Selain kata kunci, aturan kata letak kata juga dapat digunakan sebagai salah satu acuan dalam pembentukan aturan. Berdasarkan data-data yang diamati yaitu dengan mengambil kata kunci dan tata letak, sebagai dilihat pada tabel 1

Tabel 1. Aturan dan kata kunci

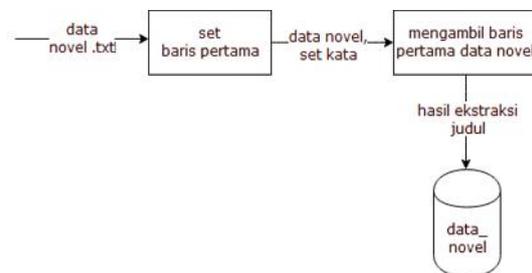
Nama	Kata Kunci	Aturan
Judul		- baris pertama dari data novel
Penulis	- Penulis - penulis - Karya - karya	- Jika kata setelah “Penulis”, “penulis”, “Karya”, “karya” maka diambil sebagai penulis - Jika kata setelah kata kunci kosong maka diambil baris berikutnya setelah kata kunci yang ditemukan
Tahun terbit	2000 sampai 2025	- Jika ditemukan rentang angka tahun 2000 - 2025 maka di ambil sebagai tahun terbit. - Menggunakan aturan rentang angka dan yang paling awal di temukan pada data novel
Penerbit	- Penerbit - penerbit - diterbit-kan - Diterbit-kan - Redaksi - redaksi	- mengambil kata setelah kata kunci yaitu “Penerbit”, “penerbit”, “Diterbitkan”, “diterbitkan”, “Redaksi”, atau “redaksi” maka di ambil sebagai penerbit.

		- Jika kata setelah kata kunci kosong maka diambil baris berikutnya setelah kata kunci
ISBN	- ISBN - isbn	- mengambil kata setelah kata kunci “ISBN” atau “isbn” maka diambil sebagai ISBN novel.
E-ISBN	- E-ISBN - e-isbn	- mengambil kata setelah kata kunci “E-ISBN” atau “EISBN” atau “e-isbn” maka diambil sebagai E-ISBN novel.

2.4 Ekstraksi Judul

Proses ekstraksi informasi (*Information Extraction* atau IE) adalah sebuah proses untuk mendapatkan informasi inti dari suatu dokumen tidak terstruktur menjadi dokumen terstruktur yang mudah dikenali. Adapun data yang terstruktur adalah data yang telah terorganisir sehingga mudah dalam suatu pencarian data sedangkan tidak terstruktur adalah data yang belum terorganisir [5].

Proses ekstraksi judul berdasarkan pengamatan data novel maka diambil kesimpulan bahwa dalam pengambilan judul novel, proses yang akan dilakukan pada penelitian ekstraksi informasi pada novel menerapkan rule-base untuk mengetahui judul dari sebuah novel adalah dengan aturan bahwa yang baris pertama dari data dokumen adalah judul dari novel dapat dilihat pada gambar 3.



Gambar 3. Ekstraksi Judul

Aturan yang diambil untuk ekstraksi informasi judul adalah menggunakan aturan yaitu mengambil baris pertama dari data novel sebagai judul. Apa yang ada di baris pertama data novel adalah judul. Contoh ekstraksi judul di lihat pada tabel 2.

Tabel 2. Ekstraksi Judul

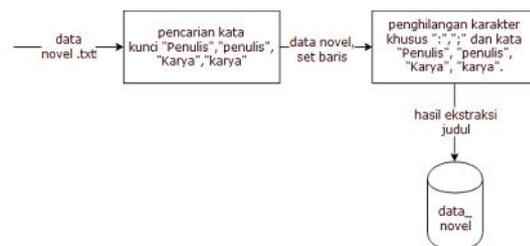
Masukan
PRIVADI DAN MARTABAT BUYA HAMKA Penulis; H. Rusydi Hamka Copyright © H. Rusydi Hamka , 2016 All rights reserve d Hat cipta dilindung undang-undang Penyunting : Laura Ariestiyanty Penyelaras Aksara: Meni Suryani & Lya Astika Penata Akssra : Abdul Waha b Ilustrasi Sampul : Robbi Gandaman a Perancan g Sampul: AS. Zuhri Digitalisasi: EHizaTittti Diterbitkan oleh Penerbit Noura (PT Mizan Publika) Anggota IKAPI Jl. Jagakarsa Raya, No. 40 Rt 007/Rw 004 Jagakarsa , Jakarta Selatan 12620 Tdp , 021-788S0556 , Fab . 021-78&S056 3 E-mail: redaksi@fnoura.mizan.com http://nourabooks.co.id ISBN 973-602-385-240-6 E-ISBN: 97S-602-385-246-8 Ebook ini didistribusikan oleh: Mizan Digital Publishing Jl. Jagakarsa Raya No. 40 Jakarta Selatan - 12620 Phone.: +62-21-7864547 (Hunting) Fax.: +62-21-7864272 email: mizandigitalpublishing@mizan.com Bahan dengan hak cipta
Keluaran
PRIVADI DAN MARTABAT BUYA HAMKA

Tabel 3. Ekstraksi Penulis

Masukan
PRIVADI DAN MARTABAT BUYA HAMKA Penulis; H. Rusydi Hamka Copyright © H. Rusydi Hamka , 2016 All rights reserve d Hat cipta dilindung undang-undang Penyunting : Laura Ariestiyanty Penyelaras Aksara: Meni Suryani & Lya Astika Penata Akssra : Abdul Waha b Ilustrasi Sampul : Robbi Gandaman a Perancan g Sampul: AS. Zuhri Digitalisasi: EHizaTittti Diterbitkan oleh Penerbit Noura (PT Mizan Publika) Anggota IKAPI Jl. Jagakarsa Raya, No. 40 Rt 007/Rw 004 Jagakarsa , Jakarta Selatan 12620 Tdp , 021-788S0556 , Fab . 021-78&S056 3 E-mail: redaksi@fnoura.mizan.com http://nourabooks.co.id ISBN 973-602-385-240-6 E-ISBN: 97S-602-385-246-8 Ebook ini didistribusikan oleh: Mizan Digital Publishing Jl. Jagakarsa Raya No. 40 Jakarta Selatan - 12620 Phone.: +62-21-7864547 (Hunting) Fax.: +62-21-7864272 email: mizandigitalpublishing@mizan.com Bahan dengan hak cipta
Keluaran
H. Rusydi Hamka

2.5 Ekstraksi Penulis

Proses ekstraksi penulis berdasarkan pengamatan data novel menerapkan rule-base pada ekstraksi penulis dilihat pada gambar 4.



Gambar 4. Ekstraksi Penulis

Aturan yang diambil untuk ekstraksi informasi penulis adalah :

1. Jika kata “Penulis atau “penulis” atau “Karya” atau “karya” maka simpan baris tersebut
2. Jika kata setelah kata kunci kosong maka diambil baris berikutnya setelah kata kunci “Penulis atau “penulis” atau “Karya” atau “karya”
3. Setelah simpan baris maka baris akan di preprocessing penghilangan karakter khusus yang ada pada baris tersebut.

Berikut contoh ekstraksi penulis di lihat pada tabel 3.

2.6 Ekstraksi Tahun Terbit Atau Terbitan Pertama

Proses ekstraksi tahun terbit atau terbitan pertama berdasarkan pengamatan data novel menerapkan rule-base pada ekstraksi tahun terbit dilihat pada gambar 5.



Gambar 5. Ekstraksi Tahun Terbit

Aturan yang diambil untuk ekstraksi informasi tahun terbit atau terbitan pertama adalah

1. Mengambil rentang angka 2001 - 2025 maka ambil sebagai tahun terbit
2. mengambil angka yang paling awal ditemukan.

Berikut contoh ekstraksi tahun terbit di lihat pada tabel 4.

Tabel 4. Ekstraksi Tahun Terbit

Masukan
PRIVADI DAN MARTABAT BUYA HAMKA Penulis; H. Rusydi Hamka Copyright © H. Rusydi Hamka , 2016 All rights reserve d Hat cipta dilindung undang-undang Penyunting : Laura Ariestiyanty Penyelaras Aksara: Meni Suryani & Lya Astika Penata Akssra : Abdul Waha b Ilustrasi Sampul : Robbi Gandaman a Perancan g Sampul: AS. Zuhri Digitalisasi: EHizaTitti Diterbitkan oleh Penerbit Noura (PT Mizan Publika) Anggota IKAPI Jl. Jagakarsa Raya, No. 40 Rt 007/Rw 004 Jagakarsa , Jakarta Selatan 12620 Tdp , 021-788S0556 , Fab . 021-78&S056 3 E-mail: redaksi@fnoura.mizan.com http://nourabooks.co.id ISBN 973-602-385-240-6 E-ISBN: 97S-602-385-246-8 Ebook ini didistribusikan oleh: Mizan Digital Publishing Jl. Jagakarsa Raya No. 40 Jakarta Selatan - 12620 Phone.: +62-21-7864547 (Hunting) Fax.: +62-21-7864272 email: mizandigitalpublishing@mizan.com Bahan dengan hak cipta
Keluaran
2016

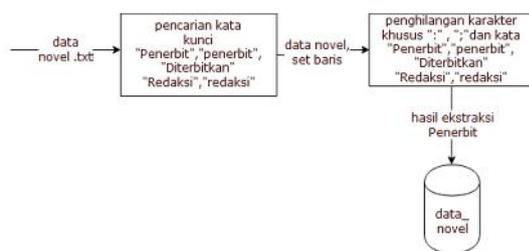
Berikut contoh ekstraksi penerbit di lihat pada tabel 5.

Tabel 5. Ekstraksi Penerbit

Masukan
PRIVADI DAN MARTABAT BUYA HAMKA Penulis; H. Rusydi Hamka Copyright © H. Rusydi Hamka , 2016 All rights reserve d Hat cipta dilindung undang-undang Penyunting : Laura Ariestiyanty Penyelaras Aksara: Meni Suryani & Lya Astika Penata Akssra : Abdul Waha b Ilustrasi Sampul : Robbi Gandaman a Perancan g Sampul: AS. Zuhri Digitalisasi: EHizaTitti Diterbitkan oleh Penerbit Noura (PT Mizan Publika) Anggota IKAPI Jl. Jagakarsa Raya, No. 40 Rt 007/Rw 004 Jagakarsa , Jakarta Selatan 12620 Tdp , 021-788S0556 , Fab . 021-78&S056 3 E-mail: redaksi@fnoura.mizan.com http://nourabooks.co.id ISBN 973-602-385-240-6 E-ISBN: 97S-602-385-246-8 Ebook ini didistribusikan oleh: Mizan Digital Publishing Jl. Jagakarsa Raya No. 40 Jakarta Selatan - 12620 Phone.: +62-21-7864547 (Hunting) Fax.: +62-21-7864272 email: mizandigitalpublishing@mizan.com Bahan dengan hak cipta
Keluaran
Noura (PT Mizan Publika)

2.7 Ekstraksi Penerbit

Proses ekstraksi penerbit berdasarkan pengamatan data novel menerapkan rule-base pada ekstraksi penerbit dilihat pada gambar 6.



Gambar 6. Ekstraksi Penerbit

Aturan yang diambil untuk ekstraksi informasi penerbit adalah

1. Menggunakan kata kunci yaitu Jika kata “Penerbit” atau kata “penerbit” atau kata “Diterbitkan oleh” atau kata “redaksi” atau kata “Redaksi” atau maka simpan baris yang terdapat kata kunci
2. Jika kata setelah kata kunci kosong maka diambil baris berikutnya setelah kata kunci
3. menghilangkan karakter khusus pada baris kata kunci

2.8 Ekstraksi ISBN

Proses ekstraksi ISBN berdasarkan pengamatan data novel menerapkan rule-base pada ekstraksi ISBN dilihat pada gambar 7.



Gambar 7. Ekstraksi ISBN

Aturan yang diambil untuk ekstraksi informasi ISBN adalah :

1. Jika kata “ISBN” atau “isbn” maka simpan baris yang terdapat kata kunci
2. Mengambil karakter angka pada baris kata kunci. Berikut contoh ekstraksi ISBN di lihat pada tabel

Tabel 6. Ekstraksi ISBN

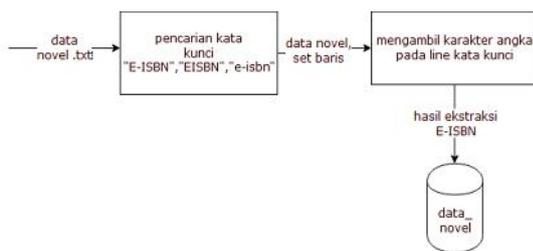
Masukan
PRIBADI DAN MARTABAT BUYA HAMKA Penulis; H. Rusydi Hamka Copyright © H. Rusydi Hamka , 2016 All rights reserve d Hat cipta dilindung undang-undang Penyunting : Laura Ariestiyanty Penyelaras Aksara: Meni Suryani & Lya Astika Penata Akssra : Abdul Wahab Ilustrasi Sampul : Robbi Gandaman a Perancan g Sampul: AS. Zuhri Digitalisasi: EHizaTitti Diterbitkan oleh Penerbit Noura (PT Mizan Publika) Anggota IKAPI Jl. Jagakarsa Raya, No. 40 Rt 007/Rw 004 Jagakarsa , Jakarta Selatan 12620 Tdp , 021-788S0556 , Fab . 021-78&S056 3 E-mail: redaksi@frnoura.mizan.com http://nourabooks.co.id ISBN 973-602-385-240-6 E-ISBN: 97S-602-385-246-8 Ebook ini didistribusikan oleh: Mizan Digital Publishing Jl. Jagakarsa Raya No. 40 Jakarta Selatan - 12620 Phone.: +62-21-7864547 (Hunting) Fax.: +62-21-7864272 email: mizandigitalpublishing@mizan.com Bahan dengan hak cipta
Keluaran
973-602-385-240-6

Tabel 7. Ekstraksi E-ISBN

Masukan
PRIBADI DAN MARTABAT BUYA HAMKA Penulis; H. Rusydi Hamka Copyright © H. Rusydi Hamka , 2016 All rights reserve d Hat cipta dilindung undang-undang Penyunting : Laura Ariestiyanty Penyelaras Aksara: Meni Suryani & Lya Astika Penata Akssra : Abdul Wahab Ilustrasi Sampul : Robbi Gandaman a Perancan g Sampul: AS. Zuhri Digitalisasi: EHizaTitti Diterbitkan oleh Penerbit Noura (PT Mizan Publika) Anggota IKAPI Jl. Jagakarsa Raya, No. 40 Rt 007/Rw 004 Jagakarsa , Jakarta Selatan 12620 Tdp , 021-788S0556 , Fab . 021-78&S056 3 E-mail: redaksi@frnoura.mizan.com http://nourabooks.co.id ISBN 973-602-385-240-6 E-ISBN: 97S-602-385-246-8 Ebook ini didistribusikan oleh: Mizan Digital Publishing Jl. Jagakarsa Raya No. 40 Jakarta Selatan - 12620 Phone.: +62-21-7864547 (Hunting) Fax.: +62-21-7864272 email: mizandigitalpublishing@mizan.com Bahan dengan hak cipta
Keluaran
97S-602-385-246-8

2.9 Ekstraksi E-ISBN

Proses ekstraksi E-ISBN berdasarkan pengamatan data novel menerapkan rule-base pada ekstraksi E-ISBN dilihat pada gambar 8



Gambar 8. Ekstraksi E-ISBN

Aturan yang diambil untuk ekstraksi informasi E-ISBN adalah

1. Jika kata "E-ISBN" atau "EISBN" atau "e-isbn" maka simpan baris yang terdapat kata kunci
 2. mengambil karakter angka pada line kata kunci
- Berikut contoh ekstraksi E-ISBN di lihat pada tabel 7.

2.10 Pengujian Ekstraksi Informasi

Pengujian ekstraksi merupakan tahap yang memiliki tujuan untuk mengetahui performa dari metode *rulebase* yang digunakan pada sistem yang dibangun. menghasilkan ekstraksi yang diinginkan baik judul, penulis, tahun terbit, penerbit, isbn, dan eisbn. Pengujian dilakukan dengan 50 data set.

a. Hasil Perbandingan Ekstraksi Informasi

Ekstraksi informasi perbandingan dokumen dengan data sistem pada data novel yaitu judul, penulis, tahun terbit atau terbitan pertama, penerbit, ISBN dan E-ISBN. Hasilnya dapat dilihat pada tabel 8.

Tabel 8. Hasil Ekstraksi

No	Judul novel setelah konversi	Hasil Ekstraksi	Kesimpulan
1	Pribadi dan martabat buya hamka	Pribadi dan martabat buya hamka	Sesuai
2	Happily Ever After	Happily Ever After	Sesuai
3	J0mb10: Sebuah Komedi	J0mb10: Sebuah Komedi	Sesuai

No	Judul novel setelah konversi	Hasil Ekstraksi	Kesimpulan	No	Judul novel setelah konversi	Hasil Ekstraksi	Kesimpulan
	Cinta	Cinta			KAMPUNG SENDIRI	KAMPUNG SENDIRI	
4	LO, TUNANGA N GU!!!	LO, TUNANGA N GU!!!	Sesuai	27	Marrmgeabfic	Marrmgeabfic	Sesuai
5	P0sesTF	P0sesTF	Sesuai	28	Let's Break Up	Let's Break Up	Sesuai
6	Memmkimii	Memmkimii	Sesuai	29	777e,7¢/La/2,;	777e,7¢/La/2,;	Sesuai
7	Move On #Crazyl,ove	Move On #Crazyl,ove	Sesuai	30	HERE. AFTER	HERE. AFTER	Sesuai
8	My Favour~i+e Dis+rc1c+ion	My Favour~i+e Dis+rc1c+ion	Sesuai	31	Cat Lover's Book	Cat Lover's Book	Sesuai
9	MIMPI BAYANG JINGGA	MIMPI BAYANG JINGGA	Sesuai	32	Memories to Forget	Memories to Forget	Sesuai
10	Jctuh Cinta Diam-Diam #2	Jctuh Cinta Diam-Diam #2	Sesuai	33	When Love Walked In	When Love Walked In	Sesuai
11	Setelah Kamu Pergi	Setelah Kamu Pergi	Sesuai	34	Apa Pun Qdain -I- \uJ',an	Apa Pun Qdain -I- \uJ',an	Sesuai
12	Spy in Love	Spy in Love	Sesuai	35	?M~*%~@<>-	?M~*%~@<>-	Sesuai
13	LDR #CrazyLove	LDR #CrazyLove	Sesuai	36	SOMEWHE RE ONLY WE KNOW	SOMEWHE RE ONLY WE KNOW	Sesuai
14	Married in Trouble	Married in Trouble	Sesuai	37	fifimclwle	fifimclwle	Sesuai
15	SWU BERSANH Wm	SWU BERSANH Wm	Sesuai	38	Unforgettabl e	Unforgettabl e	Sesuai
16	ISTRIKU SERIBU	ISTRIKU SERIBU	Sesuai	39	LAST FOREVER	LAST FOREVER	Sesuai
17	QMTA 5goAToIAu QuI	QMTA 5goAToIAu QuI	Sesuai	40	Goodbye Happiness	Goodbye Happiness	Sesuai
18	Manusia Setengah Salmon	Manusia Setengah Salmon	Sesuai	41	Péllow TQZK	Péllow TQZK	Sesuai
19	FREYA VALIKRYE	FREYA VALIKRYE	Sesuai	42	Ocean Melodq	Ocean Melodq	Sesuai
20	MUHAMM AD: Lelaki Penggengga m Hujan	MUHAMM AD: Lelaki Penggengga m Hujan	Sesuai	43	Catatan Ayah tentang Cintanya kepada Ibu	Catatan Ayah tentang Cintanya kepada Ibu	Sesuai
21	SANG PEMIMPI	SANG PEMIMPI	Sesuai	44	DROP OUT	DROP OUT	Sesuai
22	Kambingjantan	Kambingjantan	Sesuai	45	Siap Kuliah	Siap Kuliah	Sesuai
23	7 ""Yll~'~vi'R @_~!wW~ Scvlm	7 ""Yll~'~vi'R @_~!wW~ Scvlm	Sesuai	46	QEWNG /<5A ~(QEWNG /<5A ~(Sesuai
24	Look! I'm On Fire	Look! I'm On Fire	Sesuai	47	FEEL REAL	FEEL REAL	Sesuai
25	Di Peia/45¢	Di Peia/45¢	Sesuai	48	4*?""	4*?""	Sesuai
26	GBLANDA NGAN DI	GBLANDA NGAN DI	Sesuai	49	Melampaui Mimpi	Melampaui Mimpi	Sesuai
				50	Quin Affair	Quin Affair	Sesuai

b. Pengujian Akurasi

Pengujian akurasi ini dilakukan untuk mengetahui akurasi dalam mengidentifikasi hasil ekstraksi informasi. Tingkat akurasi diperoleh dengan perhitungan berikut

- Pengujian data benar : 50

$$\begin{aligned}
 & - \text{ Banyak data uji : } 50 \\
 & - \\
 & - A = \frac{p}{B} \frac{D}{a} \frac{B}{u} \times 100 \% \\
 & \qquad \qquad \qquad = \frac{5}{5} \times 100 \% \\
 & \qquad \qquad \qquad = 100 \%
 \end{aligned}$$

3. PENUTUP

Kesimpulan yang didapat dari penelitian yang telah dilakukan diketahui bahwa dari 50 data novel berhasil di ekstraksi sebanyak 100% yaitu judul, pengarang, tahun terbit, penerbit, ISBN dan E-ISBN dengan ketentuan berdasarkan hasil konversi, meskipun terdapat kesalahan hasil konversi judul karena dalam dataset terdapat jenis font yang sulit dikenali oleh sistem. Berdasarkan hal tersebut di simpulkan bahwa metode rule-base untuk penelitian ekstraksi informasi pada teks novel sangat baik

Berdasarkan hasil penelitian yang telah dilakukan, masalah yang muncul pada penelitian ini adanya hasil konversi data novel ke format .txt. Konversi dilakukan menggunakan aplikasi FreeOCR masih terdapat kesalahan pada judul novel yang tidak sesuai dengan data asli novel. Saat melakukan konversi data format JPG ke format .txt, terdapat font yang tidak dapat dikenali oleh sistem freeOCR sehingga menghasilkan konversi yang berbeda dengan data asli. Adapun saran untuk kajian lebih lanjut sebagai berikut.

1. Penambahan metode atau algoritma pendeteksian kesalahan teks pada dokumen bahasa indonesia.
2. Untuk pengembangan selanjutnya, menambahkan lebih banyak data novel untuk memperkaya aturan sehingga mendukung penanganan variasi data yang berbeda.

DAFTAR PUSTAKA

- [1] Susilawati Erna. Analisis Pengaruh kepribadian dan Harga terhadap *Impulse Buying* pada Produk Novel di Toko Buku Bandung Book Center Wilayah Bandung. *Jurnal Unikom*. 2016: 1-8.
- [2] Jiang Jing. *Information Extraction from Text*. Spinger Science and Business Media. 2012.
- [3] Peleato Ramon Aragues, Jean-Cedric Chappelier, Martin Rajman. *Using Information Extraction to Classify Newspapers Advertisements*. *Journal International on Statistical Analysis of Textual Data (JADT)*. 2000.
- [4] Riaddy Aditya Iftikar, Yuliant Sibaroni, Annisa Aditsania. Ekstraksi Informasi pada Makalah Ilmiah dengan Pendekatan *Supervised Learning*. *Jurnal Program Studi Ilmu Komputasi, Fakultas Informatika, Universitas Telkom*. 2016: 1-7.
- [5] Ismaya Agny. Algoritma Ekstraksi Informasi Berbasis Aturan. *Jurnal Nasional Teknik Elektor dan Teknik Informasi (JNTETI)*. 2014; 3(4): 242-247
- [6] M. Habibi. Ekstraksi Informasi Dengan Metode *Rule-Based* Untuk Evaluasi Pemahaman Fisika Kinematika. *Journal Informatika*. 2016.
- [7] HS Karibun M. Penerapan Metode *Rule-Based* Dengan *Unsupervised Learning* Untuk Pelabelan Dokumen Berbahasa Indonesia. *Journal Informatika*. 2005: 1-10.
- [8] Al Amin Imam Husni, Suhartono. Sistem Pakar Mendeteksi Kerusakan Rambut Menggunakan Basis Aturan (*Rule Base Reasoning*) dengan Metode *Forward Chaining*. *Jurnal Sistem Informasi Bisnis*. 2012; 3: 134-138.
- [9] Galih R Sandhika, Erik, Muhammad Lukman Hakim. Penerapan Teknik OCR (*Optical Character Recognition*) pada Aplikasi Terjemah Kitab Fiqih *Safinah An-Naja* menggunakan *Readiris*. Jurusan Teknik Informatika Universitas Pasundan Bandung. 2014.
- [10] Hulth Anette, Beata B. Megyesi. *A Study on Automatically Extracted Keywords in Text Categorization*. *Association for Computational Linguistics*. 2006: 537-544.