INFORMATION EKSTRACTION FOR TEXT DOCUMENTS NOVEL

Arbi Insani[1] , Nelly Indriani W., S.Si., M.T.[2]

[1,2] Informatics Engineering – Indonesia Computer University
Jalan Dipatiukur 112-116 Bandung
E-mail : [1] arbi_benai@yahoo.com, [2]nelly.indriani@email.unikom.ac.id

ABSTRACT

Information in digital form becomes an alternative to get accurate and fast information. In general, each library to collect information and identity of a document is done by filling the data into the system, the problem faced by the librarian is every time period to time the document each year increases and the number of resulting in the data collection of the novel many errors and takes time the old in the novel collection. Based on the above, it is necessary to build a system so that in the process of data collection can be done automatically by the system that is by extracting the information easily, quickly and produce the correct information data. The process of information extraction is a process to get core information from unstructured documents into structured documents. The extraction of information using rule-based is a method that uses factual rules from the data analyzed. For accuracy testing using 50 books in the form of novels, it shows good results, which means the of information extraction in this study can be used to extract the data identity in the novel.

Keywords : *Rule-based*, information ekstraction, novel.

1. INTRODUCTION

The process of development of information and communication technology is currently so rapid and increasing, especially information in the form of novels, journals, magazines, articles, newspapers, and news. Information in digital form becomes an alternative to get accurate and fast information. The novel is a long essay and prose and contains a series of stories of a person's life with others as well as fiction stories around him by show the character of figure and charater each other.[1]. The development of the novel always increases every time like the example novel "Manusia Setengah Salmon", "7 Secrets Shope" and other. In general, every library collect information and identity of a novel document is done by filling the data into the system, the librarian's problem is every time period to time the document each year increases and the number of resulting in the data collection of the novel many errors and requires a long time in the novel collection. Therefore it is necessary to build a system so that in the process of data collection can be done automatically by the system that is by extracting information easily, quickly and produce the correct information data.

Information ekstaraction process (*Information Extraction* or IE) is a process for finding structured information from unstructured or semi-structured documents. The structured data is data that has been organized so easily in a data search whereas unstructured data is not yet organized [2]. Examples of information extraction are extracting newspaper ads, news, scientific works and so forth according to previous research by Ramón using information extraction to classify newspaper ads [3] and research by Riaddy is to discuss the extraction of information for scientific papers using a supervised learning approach [4]. In the previous research it was concluded that the extraction of information on documents of LHP (Laporan Hasil Penelitian) can be done by using rule-based classification with accuracy of training data of 98.27% and accuracy of test data of 89.77% [5]. The research by Muhammad Habibi who performed Information Extraction for the evaluation of kinematics physics understanding can be done by rule-based method with the result of system accuracy that has been built ranging between 90.6% - 95.4% based on F - Measure testing system [6]. In a study by M. Karbun concluded that the application of rule-based methods with unsupervised learning for labeling Indonesian documents can recognize 97.85% token [7]

According the description above, it is necessary to conduct research to build a novel text document information extracting system by using rule-base method in order to collect information accurate, fast, easy and does not require long time. With the system is expected to be easier and faster in the collection of novel information.

2. CONTENT OF RESEARCH

2.1 Method

In this study used the rule-base method for extracting novel text document information. Where Rule base is also known as Hand-crafting or Knowledge engineering approach. It is called the rule base approach because it utilizes human expertise to manually create rules manually through the process of manuasi understanding itself. Human experts or experts are required to be able to understand the SEI and system domain [8].

The general overview of system analysis that will be built on the extraction of this text document information is as follows seen in Figure 1.

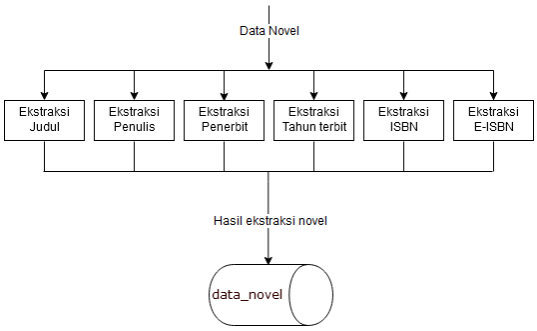


Figure 1. Illustration system

The System Stage illustration will be explained as follows.

1. In the early stages user input file text novel that .txt format to the system.
2. Next the system will ekstraction the informations like a tittle, author, publisher, year published, ISBN, E-ISBN from the novel data.
3. The last stage information from novel data will be saved in database.

2.2 Analysis of input data

Analysis of input data in this study using novel data in jpg format which is converted to .txt. using the OCR (Optical Character Recognition) application is an application that works to scan images on images and made text, and this application can also be an additional support / application for scanners. With the OCR, handwritten Image, typewriter or computer text, can be manipulated into different formats [9].

The analysis focuses on searching for keywords and rules that will be used to extract the novel data information. Here are the novel data before converting to .txt seen in Figure 2 below.

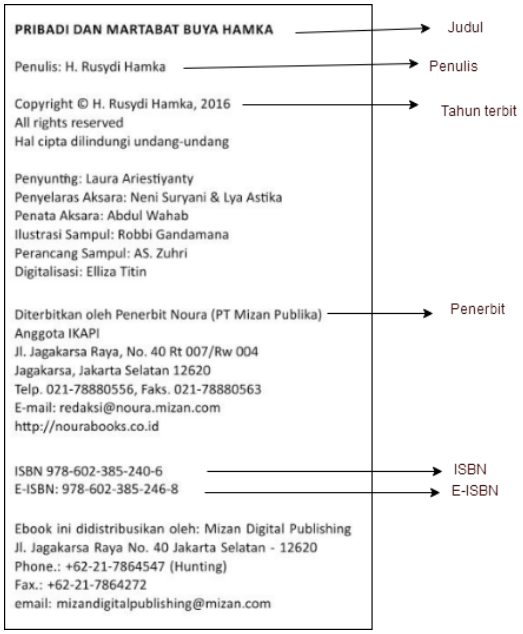


Figure 2. novel data

In the picture 2 novel data will be taken 6 data extraction information that is:

1. "PRIBADI DAN MARTABAT BUYA HAMKA" is the title of a novel that appears at the beginning of the novel data so that this can be a rule in extracting information.
2. "H. Rusydi Hamka" is the author of the novel for the retrieval of writers obtained from the keyword because before the word H. Rusydi Hamka there is a keyword that is the author.
3. "2016" is the year published from the novel for the forthcoming year using the range of the value of the numbers 2001 - 2025 because in general the rising year is always located not far from the title.
4. "Noura (PT Mizan Publika)" is the publisher of a novel for publishers retrieving keywords because before noura word there is a publisher word.
5. "973-602-385-240-6" is a novel ISBN for extracting using keywords because before 973-602-385-240-6 there is the word ISBN.
6. "978-602-385-246-8" is a novel E-ISBN for extracting using keywords because before 978-602-385-246-8 there is the word E-ISBN.

2.3 Analysis of Forming Rules and Keyword

A keyword is a word that becomes a reference of data [10]. From the novel data can be concluded to take an extraction of information on the novel need to use rule-base that is with the rules of word layout and keywords on the word. Based on the observed data that is by taking rules and keywords as seen in table 1

Table 1. Rules and keyword

Identification	Keyword	Adjective
Tittle		- Taking the title of the novel is the first line of the novel data
Author	- Penulis - penulis - Karya - karya	- Search for keywords ie If the word after "Author", "author", "Work", "works" then taken as the author - If the word after the keyword is empty then the next line is taken after the keyword is found
Year published		- If it finds the range of numbers 2000 - 2025 then taken as the year rises. - Use rules about range of numbers and the earliest found in the novel data
Publisher	- Penerbit - penerbit - diterbitkan - Diterbitkan - 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	- take the word after the keyword "ISBN" or "isbn" then taken as a novel ISBN.
E-ISBN	- E-ISBN - e-isbn	- take the word after the keyword "E-ISBN" or "EISBN" or "e-isbn" then taken as a novel E-ISBN.

2.4 Ekstraction Tittle

The information extraction process (IE) is a process for obtaining core information from an unstructured document into easily recognizable structured documents. The structured data is data that has been organized so easily in a search data whereas unstructured data is not yet organized [5].

The process of extracting the title based on the observation of the novel data then it is concluded that in the taking of the title of the novel, the process to be performed on the research of information extraction on the novel applying rule-base to find the title of a novel is by the rule that the first line of document data is the title of the novel can be seen in figure 3.

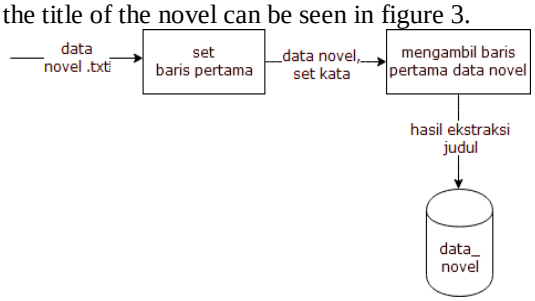


Figure 3. Ekstraction Tittle

The rule taken for the extraction of title information is to use a rule that takes the first line of the novel data as the title. what is in the first line of the novel data is the title. Examples of title extraction are seen in table 2.

Table 2. Ekstraction Tittle

Input
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 Waha b llustrasi 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@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
Output
PRIBADI DAN MARTABAT BUYA HAMKA

2.5 Ekstraction Author

The author extraction process based on observation of novel data apply rule-base on writer extraction seen on gamber 4.

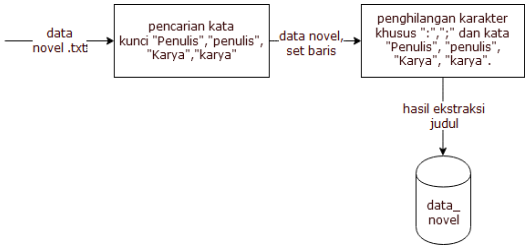


Figure 4. Ekstraction Author

The rules taken for the author's information extraction are

1. If word “Penulis” or “penulis” atau “Karya” or “karya” then save the line.
2. If word after keyword empty then take the line after keyword “Penulis” or “penulis” or “Karya” or “karya”
3. After saving the line then the line will be in preprocessing the removal of special characters that exist on that line.

The following examples of author extraction are seen in table 3

Table 3. Ekstraction Author

Input
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 Waha b llustrasi 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@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
Output
H. Rusydi Hamka

2.6 Ekstraction year published and first published

The extraction process of the year published or the first published based on the observation of novel data applies rule-base on the extraction of the rising year seen in gamber 5.

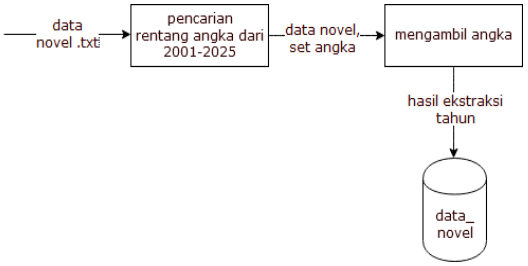


Figure 5. Ekstraction year published

Rules taken for the extraction of information published or published the first year is

1. Taking a range of numbers from 2001 - 2025 then take as the year rises.
2. Take the earliest number found.

The following examples of published year extraction are seen in table 4

Table 4. Ekstraction year published

Input
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 Waha b llustrasi 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@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
Output
2016

2.7 Ekstraction Publisher

The publisher extraction process based on the observation of novel data applies rule-base on issuer extraction seen in figure 6.

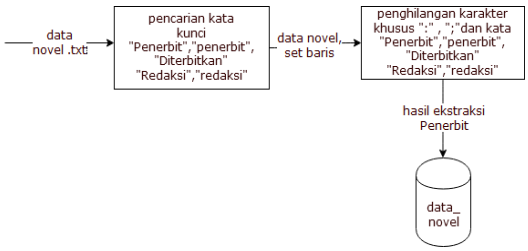


Figure 6. Ekstraction Publisher

The rules taken for publisher information extraction are

1. Use keyword if word “Penerbit” or word “penerbit” or word “Diterbitkan oleh” or word “redaksi” or word “Redaksi” or then save the existing row of keywords.
2. If the word after the keyword is empty then the next row is taken after the keyword.
3. removes special characters in the keyword row.

There an example of publisher extraction seen in table 5.

Table 5. Ekstraction Publisher

Input
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 Waha b llustrasi 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@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
Output
Noura (PT Mizan Publika)

2.8 Ekstraction ISBN

The ISBN extraction process based on observations of novel data applies rule-base on ISBN extraction seen in figure 7.

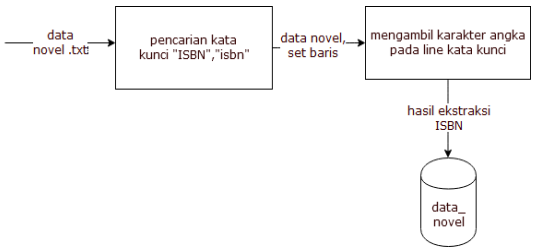


Figure 7. Ekstraction ISBN

The rules taken for the extraction of ISBN information are

1. If the word "ISBN" or "isbn" then save the existing row of keywords.
2. Takes a numeric character on the keyword row.

The following examples of ISBN extraction are seen in table 6

Table 6. Ekstraction 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 Waha b
llustrasi 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@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

Output
973-602-385-240-6

2.9 Ekstraction E-ISBN

E-ISBN extraction process based on observation of novel data apply rule base on E-ISBN extraction seen on figure 8

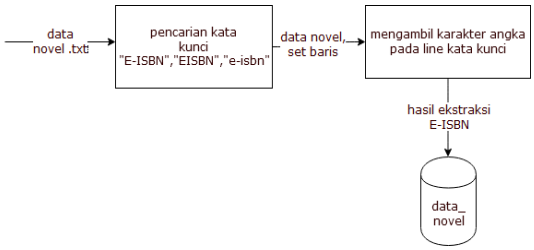


Figure 8. Ekstraction E-ISBN

The rules taken for the extraction of E-ISBN information are

1. If word “E-ISBN” or ”EISBN” or “e-isbn” then save the existing row of keywords.
2. take the number character on the keyword line.

The following examples of E-ISBN extraction are seen in table 7.

Table 7. Ekstraction E-ISBN

Input
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 Waha b
llustrasi 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@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

Output
97S-602-385-246-8

2.10 Testing Ekstraction Information

Extraction testing is a step that has a purpose to determine the performance of rulebase method used on the built system. produce the desired extraction of both title, author, year of publication, publisher, isbn, and eisbn. Testing is done with 50 data sets.

a. Result Comparison Ekstraction Information

Extraction of comparative document information with system data on novel data ie title, author, first year of publication or issue, publisher, ISBN and E-ISBN. The results can be seen in table 8.

Table 8. Result Ekstraction

No	Tittle’s novel after conversion	Result ekstraction	Conclusion
1	Pribadi dan martabat buya hamka	Pribadi dan martabat buya hamka	Suitable
2	Happily Ever After	Happily Ever After	Suitable
3	J0mbl0: Sebuah Komedi Cinta	J0mbl0: Sebuah Komedi Cinta	Suitable
4	LO, TUNANGAN GU!!!	LO, TUNANGAN GU!!!	Suitable
5	P0sesTF	P0sesTF	Suitable

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

b. Accuracy Testing

Accuracy testing is done to determine the accuracy in identifying the results of information extraction. The level of accuracy is obtained by the following calculations

- Testing true data : 50
- Amount of data : 50
- $$Akurasi = \frac{pengujian\ Data\ Benar}{Banyak\ data\ uji} x\ 100\ \%$$

$$= \frac{50}{50} x\ 100\%$$

= 100 %

3. CONCLUSION AND RECOMMENDATION

3.1 Conclusion

The conclusion obtained from the research that has been done is known that from 50 novel data is successfully extracted as much as 100% that is title, author, year published, publisher, ISBN and E-ISBN with provision based on conversion result, although there is mistake result of title conversion because in data set there are types of fonts that are difficult to recognize by the system. Based on this it is concluded that the rule-base method for research of information extraction on novel text is very good.

3.2 Recommendation

Based on the results of research that has been done, the problems that emerged in this study the result of converting novel data into .txt format. Conversions made using the FreeOCR application still have errors on novel titles that do not match the original data of the novel. When converting JPG format data to .txt format. there is a font that can not be recognized by freeOCR system so that it produces a different conversion than the original data. As for suggestions for further study as follows.

1. Added method or algorithm detection of false text on Indonesian laguage document.
2. To next development, add more novel data for survive the accuracy result.

Bibliography

[1] Susilawati Erna. Analisis Pengarh 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 Imformation 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.