



## PERINGKAS OTOMATIS TEKS BERBAHASA ARAB MENGGUNAKAN ALGORITMA *TEXTRANK*

Muhammad Fikri Hidayattullah <sup>1)</sup>, Ardhiyan Azizi <sup>2)</sup>

<sup>1</sup> Program Studi D IV Teknik Informatika, Politeknik Harapan Bersama

<sup>2</sup> Program Studi Teknik Informatika, Universitas Ahmad Dahlan

email: <sup>1</sup> fikri@poltektegal.ac.id, <sup>2</sup> ardhiyan.azeezee@gmail.com

### ARTICLE INFO

#### Article History:

Received : 27 April 2021

Revised : 14 June 2021

Accepted : 28 June 2021

Published : 30 June 2021

#### Keywords:

Automatic Summarizer

Arabic

TextRank

Artificial Intelligence

Machine Learning

#### IEEE style in citing this article:

M. F. Hidayattullah and A. Azizi, "Peringkasan Otomatis Teks Berbahasa Arab Menggunakan Algoritma TextRank",

Jurnal.ilmiah.informatika, vol. 6, no. 1, pp. 33-42, Jun. 2021.

### ABSTRACT

Increasingly, the amount of data in the form of text documents scattered on the internet is getting bigger. It took a very long time to get the information from each of these documents. For this reason, several researchers developed the Automatic Text Summarizer to summarize text automatically, so that the time needed to get important information from the entire document can be faster. Research that focuses on automatic summarization of Arabic texts is very rare. In fact, there are more than 300 million Arabic speakers in the world and Arabic is the official language at the United Nations. Therefore, this study develops a model that can perform text summarization automatically using the TextRank algorithm. The test results using Q&A Evaluation show very good results with details of the suitability of the summary results with the original text by 90%, the suitability of the summary results with Arabic grammar is 91.43%, the suitability of the summary results is 90%, the ease of understanding the summary results is 90%. and the useful aspects of the model developed were 91.43%.

© 2021 Jurnal Ilmiah Informatika (Scientific Informatics Journal) with CC BY NC licence

## 1. PENDAHULUAN

Perkembangan teknologi cerdas semakin hari semakin canggih. Teknologi tersebut diimplementasikan pada berbagai bidang kehidupan manusia. Salah satu bentuk penerapan teknologi cerdas adalah pada bidang bahasa. Muncul berbagai bidang kajian riset seperti *text mining* dan

*natural language processing*. Salah satu bentuk riset terkini di dalam bidang *text mining* yang mampu memberikan loncatan spektakuler adalah *automatic text summarization* (peringkasan teks otomatis) seperti yang dilakukan Subramanian [1]. Penelitian tersebut menghasilkan sebuah model yang mampu meringkas abstrak

dari sebuah artikel ilmiah secara otomatis. Bahkan abstrak di dalam artikel penelitian tersebut hasil *generate* otomatis dari model yang diciptakan.

Semakin hari jumlah data berupa teks yang tersebar melalui media internet semakin tinggi dan mengalami *overload*. Jumlah dokumen teks elektronik yang disimpan di seluruh dunia tidak dapat dihitung banyaknya. Perkembangan internet berperan dalam penyebarluasan artikel dan dokumen teks. Dalam survei website Netcraft pada bulan Agustus 2013, mendapat respon dari 716.822.317 situs. Angka ini meningkat sebanyak 17.998.808 situs sejak Juli 2013. Berdasarkan kecenderungan 6 (enam) bulan terakhir, Netcraft memperkirakan akan ada 1 (satu) miliar situs dalam 18 bulan ke depan. Selain itu, jumlah tersebut masih harus ditambah dengan dokumen teks elektronik yang tidak terdapat di internet. Dengan banyaknya dokumen teks yang tersedia, mengakibatkan terlalu banyak informasi yang dituntut untuk dapat diakses dengan cepat dan tidak membutuhkan banyak waktu dalam pengambilan intisari yang dibutuhkan untuk mengambil keputusan berdasarkan informasi yang tersedia. *Automatic Text Summarization* adalah salah satu solusi untuk membantu pencarian inti dari dokumen atau artikel berupa deskripsi singkat (*summary*)[2][3]. Hal ini sangat bermanfaat ketika seseorang menginginkan mendapatkan informasi inti secara cepat dari suatu tulisan yang panjang.

Ringkasan (*summary*) merupakan proses penyampaian informasi penting dari sumber teks asli [4]. Ringkasan biasanya tidak lebih dari setengah teks asli dan biasanya jauh lebih sedikit dari itu [5]. Ringkasan seringkali diperlukan untuk mempercepat kerja-kerja pembacaan dokumen dalam jumlah yang relatif banyak dan memerlukan waktu lama untuk menyelesaikan pembacaannya.

Studi penelitian yang berfokus melakukan peringkasan teks berbahasa Arab secara otomatis masih sangat jarang [6], terlebih lagi di Indonesia. Penyebabnya adalah karena kerumitan bahasa Arab tersebut, baik dari segi struktur maupun morfologi. Namun demikian, sistem peringkasan bahasa Arab sangat dibutuhkan saat ini. Ada lebih dari 300 juta penutur bahasa Arab di dunia, dan bahasa Arab adalah bahasa resmi di Perserikatan Bangsa-Bangsa [7]. Penelitian di bidang peringkasan teks berbahasa Arab menjadi sebuah tantangan yang menarik. Karena riset di bidang ini tidak sekadar meringkas dokumen menjadi lebih singkat, namun juga dokumen hasil ringkasan harus sesuai dengan kaidah gramatika bahasa Arab, terlebih jika teks tersebut memiliki harakat (tanda baca).

Beberapa penelitian seputar peringkasan teks otomatis dalam berbagai bahasa telah dibuat, seperti Eris [8] membuat peringkasan teks otomatis untuk dokumen berbahasa Indonesia. Jayashree juga membuat peringkasan teks dalam bahasa Kannada [9]. Uy membuat peringkasan teks otomatis dalam bahasa Vietnam [10], sedangkan Saggion [11] membuat peringkasan teks otomatis dalam bahasa Inggris dan memberikan *review* seputar penelitian dalam bidang *automatic text summarization*.

Salah satu algoritma yang dapat digunakan untuk meringkas dokumen adalah algoritma TextRank. Algoritma *TextRank* merupakan algoritma pemrosesan teks berbasis graf dengan model pemeringkatan [12] dan juga merupakan turunan dari algoritma Google PageRank [13]. Salah satu keunikan dari algoritma ini adalah mampu bekerja tanpa menggunakan *data training*.

Penelitian ini akan mengembangkan sebuah model peringkasan otomatis teks berbahasa Arab menggunakan algoritma TextRank. Ringkasan yang dihasilkan akan

divalidasi dengan metode pengujian *Q&A Evaluation*.

## 2. KAJIAN LITERATUR DAN PEGEMBANGAN HIPOTESIS

Eris *et. al.* [8] dalam penelitiannya yang berjudul *Penerapan Algoritma TextRank untuk Automatic Summarization pada Dokumen Berbahasa Indonesia* membuat sebuah model peringkasan teks berbahasa Indonesia berbasis *website*. Sistem menerima data berupa dokumen tunggal dan menghasilkan ringkasan ekstraktif dari dokumen tersebut. Algoritma yang digunakan adalah TextRank. Algoritma TextRank bekerja dengan membangun graf. Untuk membangun sebuah graf, sistem menerima *input* berupa dokumen teks. Graf dibangun dengan memecah dokumen menjadi kalimat-kalimat sebagai *vertex-vertex* yang saling terhubung oleh *edges*. *Edges* inilah yang merepresentasikan hubungan relasi antarkalimat. Dari hasil pengujian *Q & A evaluation*, algoritma ini mampu memberikan ringkasan dengan konten informatif hingga 82,48% untuk teks ringkasan 50% dan konten informatif 93,76% untuk teks ringkasan yang dirangkum 75%.

Silvia *et. al.* [2] mengembangkan sebuah aplikasi peringkasan teks dengan nama Aplikasi *Automatic Text Summarizer*. Penelitian ini bertujuan mengembangkan sebuah algoritma peringkasan teks dengan bertumpu kepada algoritma genetika. Algoritma yang dikembangkan didasarkan pada penilaian fitur kalimat dan algoritma genetika digunakan untuk menentukan bobot fitur kalimat. Perhitungan akurasi salah satunya menggunakan *F-Measure*. *F-measure* digunakan untuk menyatakan kualitas ringkasan sistem dengan memadukan presisi dan *recall*. Nilai *F-measure* tertinggi didapatkan saat rasio ringkasan 30%.

Saggion & Poibeau [11] melakukan penelitian seputar peringkasan teks

otomatis dalam bahasa Inggris. Penelitian ini berfokus melakukan *review* terhadap berbagai penelitian seputar *automatic text summarization* yang sudah pernah dilakukan dan yang akan dikembangkan pada masa yang akan datang. Di dalam penelitian tersebut dijelaskan mengenai lima metode peringkasan teks yang umum digunakan, yaitu *Superficial Techniques, Knowledge-based Approaches, Non-Extractive Methods, Multi-document Summarization* dan *Multilingual Summarization*. Masing-masing metode tersebut memiliki pendekatan dan cara yang berbeda dalam meringkas suatu teks atau dokumen.

Qassem *et.al.* [4] dalam penelitiannya yang berjudul *Automatic Arabic Text Summarization Based on Fuzzy Logic* berhasil melakukan peringkasan teks berbahasa Arab. Penelitian ini berupaya meningkatkan akurasi dari beberapa penelitian sebelumnya yang mengambil objek yang sama yaitu bahasa Arab. Secara umum di dalam penelitian sebelumnya terkendala dengan sulitnya struktur bahasa Arab beserta morfologinya. Algoritma yang digunakan dalam penelitian ini adalah *fuzzy logic*. Hasil akhir dari penelitian ini dievaluasi dengan metode EASC dan penggunaan *fuzzy logic* terbukti mampu meningkatkan akurasi.

## 3. METODE PENELITIAN

### 3.1 Analisis Permasalahan

Peringkasan teks berbahasa Arab tidak bisa lepas dari ilmu gramatika bahasa Arab yang terdiri dari *nahwu* dan *shorof*. Ilmu *nahwu* merupakan ilmu gramatika bahasa Arab yang mengkaji bunyi akhir suatu kata (pengharakatan akhir kata). Sedangkan ilmu *shorof* merupakan ilmu gramatika bahasa Arab yang mengkaji perubahan dari suatu kata, misalnya kata kerja *dharaba-yadhribu-dharban..* Kesalahan dalam *mentashrif* dan memberikan harakat akhir akan berakibat fatal dalam penulisan sebuah teks berbahasa Arab.

Contoh sederhana suatu frase :

كتب علي

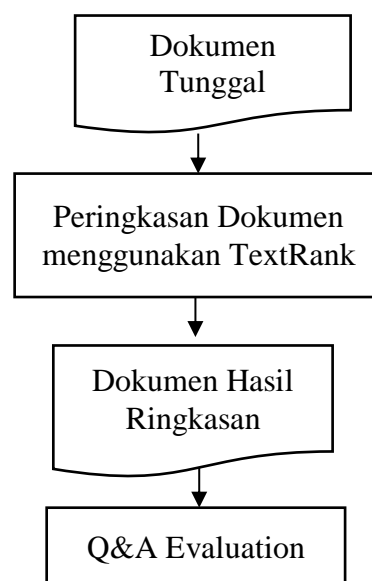
Teks di atas bisa dibaca *kataba aliyyun* (Ali telah menulis) atau juga *kutubu aliyyin* (buku-buku milik Ali). Disinilah salah satu tantangan melakukan riset di dalam peringkasan teks berbahasa Arab dengan kerumitan struktur dan morfologinya.

Pada penelitian ini akan melakukan peringkasan dokumen tunggal dengan melakukan perbandingan antara teks berharakat dan tanpa harakat. Tujuannya adalah untuk mengetahui tingkat keakuratan algoritma *TextRank* dalam melakukan peringkasan. Apakah teks yang diringkas mengalami kesalahan *tashrif* dan pengharakatan bunyi akhir suatu kata atau tidak. Oleh karena itu tahap validasi ini akan melibatkan

beberapa responden yang memiliki kemampuan ilmu gramatika bahasa Arab. Jumlah responden yang dilibatkan sebanyak 14 orang. Model pemvalidasian menggunakan *Q&A Evaluation*, yaitu dengan memberikan kuesioner kepada responden untuk dijawab terkait hasil peringkasan teks secara otomatis yang sedang dikembangkan. Hasil penilaian tersebut akan digunakan untuk menentukan sejauh mana keakuratan model peringkasan teks berbahasa Arab ini.

### 3.2 Rancangan Model

Model peringkasan teks berbahasa Arab secara otomatis akan diimplementasikan ke dalam sebuah prototipe menggunakan bahasa pemrograman Python. Dokumen yang akan diringkas hanya dokumen tunggal.



Gambar 1. Tahapan Penelitian

## 4. HASIL DAN PEMBAHASAN

### 4.1 Hasil Peringkasan

Peringkasan teks berbahasa Arab secara otomatis ini diimplementasikan ke dalam

bahasa *Python*. *Library* yang digunakan untuk mengimplementasikan algoritma *TextRank* adalah *gensim*.

```

from gensim.summarization.summarizer import summarize
from gensim.summarization import keywords
import urllib.request
from urllib.request import urlopen
from bs4 import BeautifulSoup
import requests
!pip install lxml

#mengambil teks dari file dokumen
fname = "/content/drive/My Drive/HARBER/test1.txt"
with open(fname, 'r') as myfile:
    text3=myfile.read()

print('Summary : ')
print(summarize(str(text3), word_count=100))

Requirement already satisfied: lxml in /usr/local/lib/python3.6/dist-packages (4.2.6)
Summary :
( 2 )
( تلك الكتاب لا ريب فيه هدى للمتقين )
( 2 )
وَمَا كَانَ مُحَمَّدٌ ، وَعِيسَى ، وَسُورَةُ وَمَقَاتِلُ بْنُ حَبِيبٍ ، وَزَيْدُ بْنُ أَسْلَمَ ، وَأَبُو جَرِيحٍ : أَنَّ ذَلِكَ بِمَعْنَى هَذَا ، وَالْعَرَبُ تَقَارَضُ بَيْنَ هَذَيْنِ الْإِسْمَيْنِ مِنْ أَسْمَاءِ الْإِنْتِزَاعِ فَيَسْتَعْمَلُونَ كَلِمَةً مِنْهُمَا مَكَانَ الْآخَرِ ، وَهَذَا مَشْرُوفٌ فِي كِتَابِهِمْ . وَمَعْنَى الْقَائِمِ : أَنَّ هَذَا الْكُتَيْبَ - وَهُوَ الْقُرْآنُ - لَا تَنَقَطُ فِيهِ كَلِمَةٌ إِلَّا لَزِمَ مِنْ جَدِّهِ اللَّهِ ، عَمَّا قَالَ تَعَالَى فِي الشُّعْبَةِ : ( لَمْ تَكُنْ لَكُمُ الْكُتَيْبُ لَا رَيْبَ فِيهِ مِنْ رَبِّ الْعَالَمِينَ ) [ الشُّعْبَةُ : 2 . 1 ]

```

Gambar 2. Implementasi Library Gensim

Dokumen yang diringkas ada dua jenis, yaitu dokumen berharakat dan tanpa harakat (teks Arab gundul). Salah satu teks

tanpa harakat yang akan diringkas adalah sebagai berikut :

( 2 ) ( ذلك الكتاب لا ريب فيه هدى للمتقين )

قال ابن جريج : قال ابن عباس : ذلك الكتاب : هذا الكتاب . وكذا قال مجاهد ، وعكرمة ، وسعيد بن جبیر ، والسدي ومقاتل بن حيان ، وزيد بن أسلم ، وابن جريج : أن ذلك بمعنى هذا ، والعرب تقارض بين هذين الاسمين من أسماء الإشارة فيستعملون كلا منهما مكان الآخر ، وهذا معروف في كلامهم .

و ( الكتاب ) القرآن . ومن قال : إن المراد بذلك الكتاب الإشارة إلى التوراة والإنجيل ، كما حكاه ابن جرير وغيره ، فقد أبعد النجعة وأغرق في النزاع ، وتكلف ما لا علم له به

والريب : الشك ، قال السدي عن أبي مالك ، وعن أبي صالح عن ابن عباس ، وعن مرة الهمداني عن ابن مسعود ، وعن أناس من أصحاب رسول الله صلى الله عليه وسلم : ( لا ريب فيه ) لا شك فيه

وقاله أبو الدرداء وابن عباس ومجاهد وسعيد بن جبیر وأبو مالك ونافع مولى ابن عمر وعطاء وأبو العالية والربيع بن أنس . ومقاتل بن حيان والسدي وقادة وإسماعيل بن أبي خالد . وقال ابن أبي حاتم : لا أعلم في هذا خلافا

: وقد يستعمل الريب في التهمة قال جميل [

بثينة قالت يا جميل أربنتي فقلت كلانا يا بثين مريب

: واستعمل - أيضا - في الحاجة كما قال بعضهم

[قضيئا من تهامة كل ريب وخبير ثم أجمعنا السيوفا

ومعنى الكلام : أن هذا الكتاب - وهو القرآن - لا شك فيه أنه نزل من عند الله ، كما قال تعالى في السجدة : ( ألم تنزل الكتاب لا ريب فيه من رب العالمين ) [ السجدة : 1 ، 2 ] . [ وقال بعضهم : هذا خبر ومعناه النهي ، أي : لا ترتابوا فيه ]

ومن القراء من يقف على قوله : ( لا ريب ) ويبتدئ بقوله : ( فيه هدى للمتقين ) والوقف على قوله تعالى : ( لا ريب فيه ) أولى للآية التي ذكرنا ، ولأنه يصير قوله : ( هدى ) صفة للقرآن ، وذلك أبلغ من كون : ( فيه هدى ) .

. و ( هدى ) يحتمل من حيث العربية أن يكون مرفوعا على النعت ، ومنصوبا على الحال

Gambar 3. Teks Asal

Hasil ringkasannya :

( ( 2 ) ) ذلك الكتاب لا ريب فيه هدى للمتقين )  
 وكذا قال مجاهد ، وعكرمة ، وسعيد بن جبیر ، والسدي ومقاتل بن حیان ، وزید بن أسلم ، وابن جریج : أن ذلك بمعنى هذا ،  
 . والعرب تفارض بين هذين الاسمين من أسماء الإشارة فيستعملون كلا منهما مكان الآخر ، وهذا معروف في كلامهم  
 ومعنى الكلام : أن هذا الكتاب - وهو القرآن - لا شك فيه أنه نزل من عند الله ، كما قال تعالى في السجدة : ( الم تنزيل الكتاب  
 . لا ريب فيه من رب العالمين ) [ السجدة : 1 ، 2 ]

Gambar 4. Teks Hasil Ringkasan

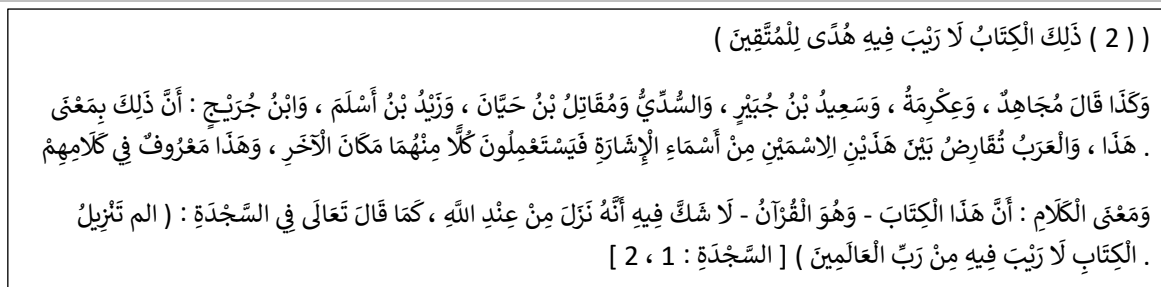
Agar dapat diketahui validitas keabsahan dari sisi gramatika bahasa Arab, maka diperlukan dokumen yang

sama namun memiliki harakat. Dokumen berharakat tersebut akan dibandingkan dengan yang tanpa harakat.

( ( 2 ) ) ذَلِكَ الْكِتَابُ لَا رَيْبَ فِيهِ هُدًى لِّلْمُتَّقِينَ )  
 قَالَ ابْنُ جُرَيْجٍ : قَالَ ابْنُ عَبَّاسٍ : ذَلِكَ الْكِتَابُ : هَذَا الْكِتَابُ . وَكَذَا قَالَ مُجَاهِدٌ ، وَعَكْرَمَةُ ، وَسَعِيدُ بْنُ جُبَيْرٍ ، وَالسُّدِّيُّ وَمُقَاتِلُ بْنُ حَيَّانَ ، وَزَيْدُ بْنُ أَسْلَمَ ، وَابْنُ جُرَيْجٍ : أَنَّ ذَلِكَ بِمَعْنَى هَذَا ، وَالْعَرَبُ تُفَارِضُ بَيْنَ هَذَيْنِ الْأَسْمَاءِ الْإِشَارَةِ . فَيَسْتَعْمَلُونَ كِلَا مِنْهُمَا مَكَانَ الْآخَرِ ، وَهَذَا مَعْرُوفٌ فِي كَلَامِهِمْ .  
 وَ ( الْكِتَابُ ) الْقُرْآنُ . وَمَنْ قَالَ : إِنَّ الْمُرَادَ بِذَلِكَ الْكِتَابِ الْإِشَارَةُ إِلَى التَّوْرَةِ وَالْإِنْجِيلِ ، كَمَا حَكَاهُ ابْنُ جُرَيْرٍ وَعَبْدُ اللَّهِ ، فَقَدْ أَبْعَدَ النَّجْعَةَ وَأَعْرَقَ فِي النَّرْعِ ، وَتَكَفَّفَ مَا لَا عِلْمَ لَهُ بِهِ .  
 وَالرَّيْبُ : الشَّكُّ ، قَالَ السُّدِّيُّ عَنْ أَبِي مَالِكٍ ، وَعَنْ أَبِي صَالِحٍ عَنِ ابْنِ عَبَّاسٍ ، وَعَنْ مَرَّةَ الْأَهْمَدَانِيِّ عَنِ ابْنِ مَسْعُودٍ ، وَعَنْ أَنَسٍ مِنْ أَصْحَابِ رَسُولِ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ : ( لَا رَيْبَ فِيهِ ) لَا شَكَّ فِيهِ .  
 وَقَالَ أَبُو الدَّرْدَاءِ وَابْنُ عَبَّاسٍ وَمُجَاهِدٌ وَسَعِيدُ بْنُ جُبَيْرٍ وَأَبُو مَالِكٍ وَنَافِعٌ مَوْلَى ابْنِ عُمَرَ وَعَطَاءٌ وَأَبُو الْعَالِيَةِ وَالرَّبِيعُ بْنُ أَنَسٍ . وَمُقَاتِلُ بْنُ حَيَّانَ وَالسُّدِّيُّ وَقَتَادَةُ وَإِسْمَاعِيلُ بْنُ أَبِي خَالِدٍ . وَقَالَ ابْنُ أَبِي حَاتِمٍ : لَا أَعْلَمُ فِي هَذَا خِلَافًا .  
 : وَقَدْ يُسْتَعْمَلُ الرَّيْبُ فِي النَّهْمَةِ قَالَ جَمِيلٌ [   
 بُنْيَنَةٌ قَالَتْ يَا جَمِيلُ أَرَبْتَنِي فَقُلْتُ كِلَانَا يَا بَنِيَّ مُرِيبُ  
 : وَأَسْتَعْمَلُ - أَيْضًا - فِي الْحَاجَةِ كَمَا قَالَ بَعْضُهُمْ  
 [فَضِينَا مِنْ نَهَامَةِ كُلِّ رَيْبٍ وَخَبِيرٌ ثُمَّ أَجْمَعْنَا السُّيُوفَا  
 وَمَعْنَى الْكَلَامِ : أَنَّ هَذَا الْكِتَابَ - وَهُوَ الْقُرْآنُ - لَا شَكَّ فِيهِ أَنَّهُ نَزَلَ مِنْ عِنْدِ اللَّهِ ، كَمَا قَالَ تَعَالَى فِي السَّجْدَةِ : ( الم تنزيل  
 الْكِتَابِ لَا رَيْبَ فِيهِ مِنْ رَبِّ الْعَالَمِينَ ) [ السَّجْدَةِ : 1 ، 2 ] . [ وَقَالَ بَعْضُهُمْ : هَذَا خَبَرٌ وَمَعْنَاهُ النَّهْيُ ، أَي : لَا تَرْتَابُوا فِيهِ ]  
 وَمَنْ الْقُرْآنَ مَنْ يَقِفُ عَلَى قَوْلِهِ : ( لَا رَيْبَ ) وَيَبْتَدِئُ بِقَوْلِهِ : ( فِيهِ هُدًى لِّلْمُتَّقِينَ ) وَالْوَقْفُ عَلَى قَوْلِهِ تَعَالَى : ( لَا رَيْبَ فِيهِ )  
 . أُولَى لِلْأَيَّةِ الَّتِي دَكَرْنَا ، وَلِأَنَّهُ يَصِيرُ قَوْلُهُ : ( هُدًى ) صِفَةً لِلْقُرْآنِ ، وَذَلِكَ أُبْلَغُ مِنْ كَوْنِ : ( فِيهِ هُدًى )  
 وَ ( هُدًى ) يَحْتَمِلُ مِنْ حَيْثُ الْعَرَبِيَّةُ أَنْ يَكُونَ مَرْفُوعًا عَلَى النَّعْتِ ، وَمَنْصُوبًا عَلَى الْحَالِ

Gambar 5. Teks Asal dengan Harkat

Hasil ringkasannya :



Gambar 6. Teks Hasil Ringkasan

Teks hasil ringkasan dari dokumen berharakat dan tanpa harakat tidak ada bedanya dari sisi kata penyusunnya. Hasil kedua ringkasan ini akan dinilai oleh para responden.

#### 4.2 Hasil Pengujian

Hasil peringkasan teks otomatis yang sudah dihasilkan perlu untuk diuji untuk mengetahui sejauh mana tingkat keakuratannya. Pengujian dilakukan dengan menggunakan metode Q&A

*Evaluation* berbentuk kuesioner skala *Likert*. Kuesioner tersebut diberikan kepada para responden yang memiliki kemampuan penguasaan gramatika bahasa Arab (nahwu-shorof). Para responden diberikan dokumen asli dan dokumen hasil ringkasan. Setelah itu, para responden diminta memberikan penilaian sesuai butir-butir pertanyaan yang ada di kuesioner.

Tabel 1. Kuesioner Penilaian

No.	Pertanyaan	1	2	3	4	5
1.	Apakah hasil ringkasan sesuai dengan teks aslinya?					
2.	Apakah hasil ringkasan sudah sesuai dengan kaidah gramatika bahasa Arab?					
3.	Apakah pengharakatan tanda baca di hasil ringkasan sudah benar semua?					
4.	Apakah hasil ringkasan mudah dipahami?					
5.	Apakah peringkasan otomatis teks berbahasa Arab ini bermanfaat?					

Jumlah responden yang diminta untuk menjawab pertanyaan-pertanyaan di kuesioner sebanyak 14 orang. Adapun rumus perhitungan skala *Likert* :

$$T \times P_n$$

T = Total jumlah responden yang memilih  
P<sub>n</sub> = Pilihan angka skor *Likert*

Setelah diperoleh nilai total dari tiap jawaban, maka langkah berikutnya adalah

melakukan interpretasi skor perhitungan. Agar mendapatkan hasil interpretasi, terlebih dahulu harus diketahui skor tertinggi (X) dan skor terendah (Y) untuk item penilaian dengan rumus sebagai berikut :

X = skor tertinggi *likert* x jumlah responden

Y = skor terendah *likert* x jumlah responden

Jumlah skor tertinggi untuk jawaban "Sangat Bagus" adalah  $5 \times 14 = 70$ . Sedangkan skor terendah untuk jawaban

"Sangat Kurang Bagus" adalah  $1 \times 14 = 14$ . Adapun tabel skala perankingannya sebagai berikut :

Tabel 2. Skala Penilaian

Total Nilai	Skala
57 - 70	Sangat Bagus
43 - 56	Bagus
29 - 42	Cukup
15 - 28	Kurang Bagus
0 - 14	Sangat Kurang Bagus

### Perhitungan Nilai Skala *Likert*

a. Skor untuk pertanyaan pertama "Apakah hasil ringkasan sesuai dengan teks aslinya?"

Berikut jawaban dari masing-masing responden :

- Responden yang memberikan skor 5 sejumlah 8  $\rightarrow (5 \times 8) = 40$
- Responden yang memberikan skor 4 sejumlah 5  $\rightarrow (4 \times 5) = 20$
- Responden yang memberikan skor 3 sejumlah 1  $\rightarrow (3 \times 1) = 3$
- Responden yang memberikan skor 2 sejumlah 0  $\rightarrow 0$
- Responden yang memberikan skor 1 sejumlah 0  $\rightarrow 0$
- Semua hasil dijumlahkan =  $40 + 20 + 3 = 63$  (**Sangat Bagus**)

Perhitungan nilai persentasenya :  
 $63/70 \times 100 = 90 \%$

b. Skor untuk pertanyaan pertama "Apakah hasil ringkasan sudah sesuai dengan kaidah gramatika bahasa Arab?"

Berikut jawaban dari masing-masing responden :

- Responden yang memberikan skor 5 sejumlah 9  $\rightarrow (5 \times 9) = 45$
- Responden yang memberikan skor 4 sejumlah 4  $\rightarrow (4 \times 4) = 16$
- Responden yang memberikan skor 3 sejumlah 1  $\rightarrow (3 \times 1) = 3$
- Responden yang memberikan skor 2 sejumlah 0  $\rightarrow 0$

- Responden yang memberikan skor 1 sejumlah 0  $\rightarrow 0$
- Semua hasil dijumlahkan =  $45 + 16 + 3 = 64$  (**Sangat Bagus**)

Perhitungan nilai persentasenya :  
 $64/70 \times 100 = 91,43 \%$

c. Skor untuk pertanyaan pertama "Apakah pengharakatan tanda baca di hasil ringkasan sudah benar semua?"

Berikut jawaban dari masing-masing responden :

- Responden yang memberikan skor 5 sejumlah 9  $\rightarrow (5 \times 9) = 45$
- Responden yang memberikan skor 4 sejumlah 4  $\rightarrow (4 \times 4) = 16$
- Responden yang memberikan skor 3 sejumlah 0  $\rightarrow 0$
- Responden yang memberikan skor 2 sejumlah 1  $\rightarrow (2 \times 1) = 2$
- Responden yang memberikan skor 1 sejumlah 0  $\rightarrow 0$
- Semua hasil dijumlahkan =  $45 + 16 + 2 = 63$  (**Sangat Bagus**)

Perhitungan nilai persentasenya :  
 $63/70 \times 100 = 90 \%$

d. Skor untuk pertanyaan pertama "Apakah hasil ringkasan mudah dipahami?"

Berikut jawaban dari masing-masing responden :

- Responden yang memberikan skor 5 sejumlah 8  $\rightarrow (5 \times 8) = 40$



- Responden yang memberikan skor 4 sejumlah 5  $\rightarrow (4 \times 5) = 20$
- Responden yang memberikan skor 3 sejumlah 1  $\rightarrow (3 \times 1) = 3$
- Responden yang memberikan skor 2 sejumlah 0  $\rightarrow 0$
- Responden yang memberikan skor 1 sejumlah 0  $\rightarrow 0$
- Semua hasil dijumlahkan =  $40 + 20 + 3 = 63$  (**Sangat Bagus**)

Perhitungan nilai presentasinya :

$$63/70 \times 100 = 90 \%$$

- e. Skor untuk pertanyaan pertama "Apakah peringkas otomatis teks berbahasa Arab ini bermanfaat?"

Berikut jawaban dari masing-masing responden :

- Responden yang memberikan skor 5 sejumlah 8  $\rightarrow (5 \times 8) = 40$
- Responden yang memberikan skor 4 sejumlah 6  $\rightarrow (4 \times 6) = 24$
- Responden yang memberikan skor 3 sejumlah 0  $\rightarrow 0$
- Responden yang memberikan skor 2 sejumlah 0  $\rightarrow 0$
- Responden yang memberikan skor 1 sejumlah 0  $\rightarrow 0$
- Semua hasil dijumlahkan =  $40 + 24 = 64$  (**Sangat Bagus**)

Perhitungan nilai presentasinya :

$$64/70 \times 100 = 91,43 \%$$

## 5. KESIMPULAN

Algoritma *TextRank* terbukti dapat digunakan untuk melakukan peringkasan dokumen berbahasa Arab. Berdasarkan pengujian *Q&A Evaluation* diperoleh hasil yang sangat bagus dengan rincian kesesuaian hasil ringkasan dengan teks aslinya sebesar 90%, kesesuaian hasil ringkasan dengan gramatika bahasa Arab sebesar 91,43%, kesesuaian pengharakatan hasil ringkasan sebesar 90%, kemudahan memahami hasil ringkasan sebesar 90% dan aspek kebermanfaatan model yang

dikembangkan sebesar 91,43%. Peringkas otomatis ini masih berbentuk prototipe sederhana, perlu dikembangkan antarmuka untuk memudahkan pengoperasian dan juga metode pengujian perlu ditambahkan *recall*, *precision* dan perhitungan akurasi.

## 6. REFERENSI

- [1] S. Subramanian, R. Li, J. Pilault, and C. Pal, "On Extractive and Abstractive Neural Document Summarization with Transformer Language Models", *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 9308-9319, Nov. 2020.
- [2] Silvia, P. Rukmana, and V. R. Aprilia, "Aplikasi Automatic Text Summarizer," *Comput. Sci. Dep. Bina Nusantara*, vol. 1969, p. 7, 2010.
- [3] I. S. Pratama, G. Alam, and Tinaliah, "Penerapan Algoritma Centroid-Based Summarization untuk Sistem Peringkasan Dokumen Berbahasa Indonesia," no. x, 2015.
- [4] L. Al Qassem, D. Wang, H. Barada, A. Al-Rubaie, and N. Almoosa, "Automatic Arabic Text Summarization Based on Fuzzy Logic," *Proc. 3rd Int. Conf. Nat. Lang. Speech Process.*, pp. 42-48, 2019.
- [5] F. Kiyani and O. Tas, "A survey automatic text summarization," *Pressacademia*, vol. 5, no. 1, pp. 205-213, 2017.
- [6] L. M. Al Qassem, D. Wang, Z. Al Mahmoud, H. Barada, A. Al-Rubaie, and N. I. Almoosa, "Automatic Arabic Summarization: A survey of methodologies and systems," *Procedia Comput. Sci.*, vol. 117, pp. 10-18, 2017.
- [7] A. B. Al-Saleh and M. E. B. Menai, "Automatic Arabic text summarization: a survey," *Artif. Intell. Rev.*, vol. 45, no. 2, pp. 203-234, 2016.
- [8] Eris, V. C. M, and J. Pragantha,

- “Penerapan Algoritma TextRank Untuk Automatic Summarization Pada Dokumen Berbahasa Indonesia,” *J. Ilmu Tek. dan Komput.*, vol. 1, no. 1, pp. 71–78, 2017.
- [9] R. Jayashree, S. Murthy K., and B. S. Anami, “Categorized Text Document Summarization in the Kannada Language by sentence ranking,” *Int. Conf. Intell. Syst. Des. Appl. ISDA*, pp. 776–781, 2012.
- [10] N. Q. Uy, P. T. Anh, T. C. Doan, and N. X. Hoai, “A study on the use of genetic programming for automatic text summarization,” *Proc. - 4th Int. Conf. Knowl. Syst. Eng. KSE 2012*, pp. 93–98, 2012.
- [11] H. Saggion *et al.*, “Automatic Text Summarization: Past, Present and Future,” 2016.
- [12] R. Mihalcea and P. Tarau, “TextRank: Bringing Order into Texts,” 2004.
- [13] R. Elbarougy, G. Behery, and A. El, “Extractive Arabic Text Summarization Using Modified PageRank Algorithm,” *Egypt. Informatics J.*, vol. 21, no. 2, pp. 73–81, 2020.