



***Multi-document Summarization Untuk Pengembangan Smart Content Pada
Kecerdasan Artifisial di Bidang Pendidikan***

PROPOSAL DISERTASI
SEMINAR BIDANG KAJIAN

NURASIAH
NIM : 99219030

PROGRAM DOKTOR TEKNOLOGI INFORMASI
UNIVERSITAS GUNADARMA

2021

Daftar Isi

1. Pendahuluan	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Kontribusi dan Manfaat.....	3
2. Tinjauan Pustaka	4
2.1 Peringkasan Teks Otomatis atau Automatic text summarization.....	4
2.1.1 Topik Penelitian Peringkasan Teks.....	5
2.1.2 Preprocessing.....	6
2.1.3 Features.....	6
2.1.4 Teknik Text Summarization.....	7
2.1.5 Metode text Summarization.....	8
2.1.6 Evaluasi Pada Text Summarization.....	8
2.1.7 Masalah Terkait Text Summarization.....	9
2.2 Perkembangan Penelitian Terkait.....	9
3. Metodologi	15
3.1 Objek Penelitian.....	15
3.2 Tahapan penelitian.....	15
3.2.1 Input Dokumen.....	16
3.2.2 Metode Concatenation.....	16
3.2.3 Preprocessing.....	16
3.2.4 Metode Peringkasan Teks.....	16
3.2.5 Metric Evaluasi.....	16
3.3 Rencana Kerja.....	17
Daftar Pustaka	18

Bab 1

Pendahuluan

1.1 Latar Belakang

Artificial intelligence atau kecerdasan buatan (AI) merupakan teknologi ilmu baru yang digunakan untuk mensimulasikan kecerdasan manusia. Kemunculan AI adalah untuk memfasilitasi kehidupan manusia. Pemanfaatan kecerdasan artificial antara lain dibidang Pendidikan, yaitu AI dapat meningkatkan proses pembelajaran. Implementasi AI di dunia pendidikan yaitu merancang sistem sedemikian rupa agar dapat membantu proses belajar mengajar [1]. Salah satu kelebihan menggunakan AI di dunia pendidikan yaitu berpotensi untuk menciptakan *personalized learnig*. AI dapat memodelkan kurikulum secara unik dan didistribusikan kepada siswa berdasarkan kemampuan personal sehingga proses pembelajaran lebih efisien dan fleksibel. *Intelligent Touring System (ITS)* adalah salah satu tipe AI modern yang dapat membantu untuk mencapai tujuan *personalized learning* [2].

Penerapan AI di bidang pendidikan seperti guru virtual dapat dijadikan sebagai sampel *learning content* untuk mengetahui perilaku siswa ketika melakukan pembelajaran secara virtual. Siswa memungkinkan untuk berinterkasi dengan guru virtual dengan cara berkomunikasi kepada sistem selayaknya seperti berkomunikasi kepada guru secara nyata [3]. Mentor virtual merupakan mesin kecerdasan artifisial yang dapat membimbing siswa dalam pembelajaran. Terdapat beberapa tutor belajar berbasis kecerdasan artifisial untuk membantu siswa belajar menulis, matematika dasar dan mata pelajaran lainnya [4].

Smart content merupakan kecerdasan artifisial yang dapat menyajikan konten sesuai dengan preferensi personal. Implementasi *smart content* di dalam pembelajaran yaitu kecerdasan artifisial dapat memberikan rekomendasi buku dan konten lain yang relevan dengan pencarian. Sebagai teknologi berbasis kecerdasan, *smart content* diperlukan untuk mendukung proses pembelajaran secara cepat [5]. Peringkasan teks dokumen atau *automatic text summarization* adalah salah satu penerapan kecerdasan artifisial untuk memudahkan pencarian konten. Pada bidang pendidikan, mesin kecerdasan artifisial dapat melakukan peringkasan dokumen sehingga dapat membantu siswa dalam memperoleh informasi secara cepat.

Peringkasan otomatis atau *automatic text summarization* merupakan cara untuk mengambil intisari informasi dari sebuah atau banyak dokumen teks. Penerapan mesin peringkasan dapat ditemukan pada berbagai bidang, seperti peringkasan artikel baik berita

maupun ilmiah, teks pendek untuk mesin pencarian, sinopsis buku, publikasi ilmiah, *headline* berita, peringkasan email, *lecture feedback* dan *review* produk. Pada bidang kesehatan, peringkasan dokumen juga diterapkan agar pembacaan dokumen medis menjadi lebih cepat dan efisien. Peringkasan teks dokumen otomatis atau *automatic text summarization* dapat menghasilkan dokumen menjadi lebih pendek dari dokumen aslinya tanpa menghilangkan inti dari isinya secara otomatis dengan berbagai algoritma [6]. Berdasarkan jumlah input dokumen, peringkasan teks dokumen dapat dibagi menjadi peringkasan dokumen tunggal atau *single document summarization* dan peringkasan multi-dokumen atau *multi-document summarization*. Peringkasan dokumen tunggal bertujuan untuk membentuk ringkasan dari hanya satu dokumen sedangkan peringkasan multi-dokumen bertujuan untuk menghasilkan ringkasan singkat dan informatif dari seluruh kumpulan dokumen.

Peringkasan multi-dokumen telah menerima banyak perhatian dari berbagai industri. *Intelligent multilingual news reporter bot* yang diberi nama Xiaomingbot telah berhasil dikembangkan untuk membuat berita. Bot ini mampu merangkum beberapa berita menjadi satu artikel kemudian menerjemahkannya ke dalam berbagai bahasa. Persyaratan pengembangan aplikasi dan data online yang berkembang pesat mendorong pengembangan *multi-document summarization* [7].

Peneliti Garmastewira, dkk [8] melakukan penelitian peringkasan multi-dokumen artikel berita bahasa Indonesia dengan *graph convolutional network* (GCN). Hasil penelitiannya menyatakan bahwa model peringkasan yang diusulkan memiliki performa yang baik. Saran penelitiannya yaitu untuk menghapus kata – kata yang tidak perlu sehingga sistem dapat mengambil kalimat yang lebih terkompresi sebagai bagian dari kalimat. Tahap *preprocessing* yang dilakukan hanya *sentence segmentasi* dan *tokenizing*.

Natural language terdiri dari kosa kata dan frase yang saling keterhubungan sangat kuat, graph dapat mewakili keterhubungan antara mereka daripada disusun secara berurutan. *Graph neural networks* (GNN) merupakan arsitektur yang ideal untuk pemrosesan natural language karena mampu memodelkan keterhubungan yang kuat antara entitas secara semantik dan sintaksis. *Graph convolutional networks* dan *graph attention networks* merupakan yang paling sering diadopsi dari GNN karena efisiensi dan kesederhanaanya untuk berintegrasi dengan jaringan neural lainnya [9].

Multi-document summarization untuk pengembangan *smart content* pada kecerdasan artificial di bidang Pendidikan mempengaruhi dalam pencarian informasi, sehingga pada penelitian ini perlu dilakukan pengembangan. Berdasarkan penelitian sebelumnya terkait

peringkasan multi-dokumen bahasa Indonesia maka pengembangan yang akan dilakukan adalah model *graph neural networks*.

1.2 Batasan Masalah

Penelitian ini berusaha mengembangkan model multi-document summarization untuk pengembangan *smart content* pada kecerdasan srtifisial di bidang Pendidikan sehingga penelitian difokuskan pada :

1. Data yang digunakan adalah dokumen yang menggunakan Bahasa Indonesia.
2. Penelitian menggunakan data *ebook* mata kuliah.
3. Proses pembelajaran yang dilakukan adalah *supervised learning*.

1.3 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah :

1. Bagaimana mengembangkan metode *multi-document summarization* untuk pengembangan *smart content* pada kecerdasan artificial di bidang Pendidikan.
2. Bagaimana mengembangkan metode klasifikasi untuk *multi-document summarization*

1.4 Tujuan

Secara umum tujuan penelitian adalah mengembangkan model *automation summarization document*. Adapun tujuan penelitian ini adalah :

1. Mengembangkan metode *multi-document summarization* untuk pengembangan *smart content* pada kecerdasan artificial di bidang Pendidikan.
2. Mengembangkan metode klasifikasi untuk *multi-document summarization*.

1.5 Kontribusi dan Manfaat

1. Pada bidang teknologi informasi hasil penelitian dapat digunakan untuk pengembangan aplikasi *smart content* kecerdasan artificial di bidang Pendidikan.
2. Pada bidang Ilmu Pengetahuan penelitian ini memberikan kontribusi untuk penelitian *Natural Language Processing* (NLP) khususnya *multi-document summarization* yang dapat dijadikan acuan untuk pengembangan selanjutnya.
3. Bagi masyarakat khususnya pelajar, penelitian ini diharapkan menjadi solusi untuk mendapatkan informasi dan rekomendasi buku agar proses pembelajaran menjadi lebih cepat.

Bab 2

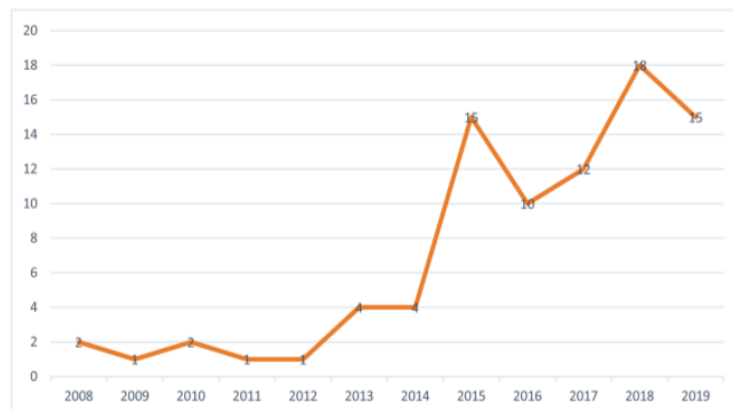
Tinjauan Pustaka

Bab ini menguraikan studi literature mengenai penelitian peringkasan teks otomatis dan perkembangan penelitian terkait peringkasan teks otomatis khususnya untuk penelitian *multi-summarization document*.

2.1 Peringkasan Teks Otomatis atau Automatic text summarization

Metode peringkasan teks dokumen dibagi menjadi dua kategori, yaitu secara ekstraktif dan abstraktif. Metode ringkasan ekstraktif memilih cuplikan yang dominan dari dokumen aslinya untuk membentuk ringkasan yang informatif. Metode ini memiliki dua komponen utama yaitu sentence ranking dan sentence selection. Metode ekstraktif memastikan ringkasan yang dihasilkan secara semantik mirip dengan dokumen aslinya [10]. Metode ringkasan abstraktif menyajikan informasi utama dari dokumen input dan anak menghasilkan ringkasan yang ringkas dan koheren. Metode abstraktif memungkinkan model menghasilkan kata dan kalimat baru dari kumpulan corpus. Metode abstraktif memiliki tantangan tersendiri karena untuk menghasilkan ringkasan membutuhkan pemahaman *natural language* yang mumpuni dan teknik generasi seperti paraphrase dan teknik *sentence fusion* [11].

Peneliti Adhika Pramita Widyassari, dkk [12] melakukan penelitian studi literatur mengenai teknik dan metode peringkasan teks dokumen otomatis. Studi ini dilakukan dengan melakukan tinjauan secara luas dan sistematis mengenai penelitian peringkasan teks otomatis. Diketahui terdapat 85 artikel ilmiah terpublikasi yang melakukan penelitian peringkasan teks dari tahun 2008 sampai dengan tahun 2019.

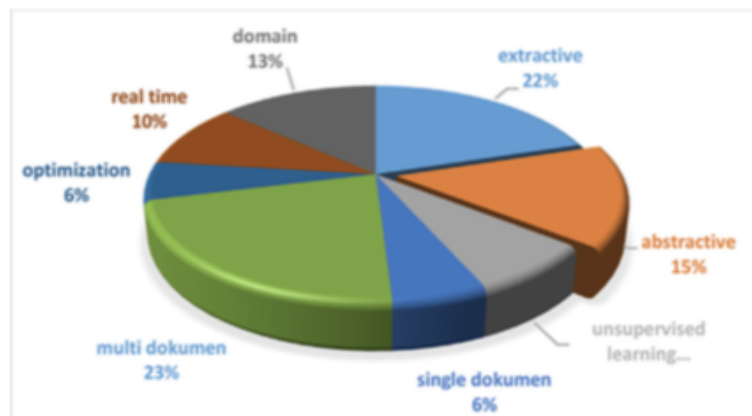


Gambar 1. Grafik Perkembangan Publikasi Artikel Ilmiah[7]

Gambar 1 menunjukkan perkembangan jumlah artikel ilmiah terpublikasi dari tahun ke tahun selama hampir 10 tahun terakhir. Hal ini menunjukkan bahwa penelitian peringkasan teks masih sangat relevan.

2.1.1 Topik Penelitian Peringkasan Teks

Penelitian terkait peringkasan teks memiliki beberapa topik dan dari 10 tahun terakhir terdapat 8 topik penelitian peringkasan teks yang menjadi tren yaitu ekstraktif, abstraktif, dokumen tunggal, multi dokumen, *optimization*, *domain* dan *real time summarization*.



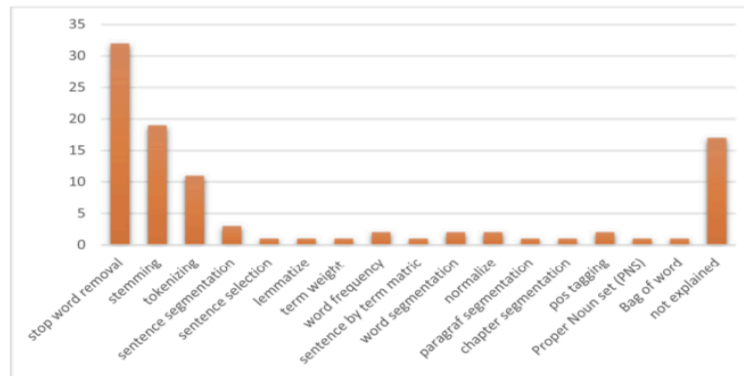
Gambae 2. Tren Topik Penelitian Peringkasan Teks[12]

Favorit topik penelitian ringkasan teks selama 10 tahun terkahir adalah *multi-document summarization*. Topik *multi-document* menjadi topik yang lebih populer karena memiliki keunikan tersendiri disbanding dengan *single document*. Hal ini juga karena *multi-document* memiliki ruang lingkup yang besar dan memiliki konten berbeda disetiap dokumen. Selain itu, *multi-document* dapat memberikan informasi dari apa yang ada diberbagai media online dan sebanding dengan banyaknya data yang ada di internet saat ini.

Favorit topik berikutnya adalah *extractive text summarization* karena lebih objektif dan tidak hanya menyajikan sudut pandang saja. Peringkasan teks secara ekstraktif memilih kata, kalimat dan paragraf yang paling penting untuk menghasilkan ringkasan. Peringkasan teks ekstraktif memiliki kendala lebih mudah dibanding dengan peringkasan teks abstraktif yang lebih kompleks. Meski begitu masih banyak persoalan yang harus dilakukan oleh para peneliti mengenai peringkasan teks ekstraktif seperti menentukan tahapan preprocessing yang sesuai dengan dataset, memilih fitur yang tepat, memaksimalkan fitur untuk menghasilkan ringkasan yang lebih baik, menentukan metode yang tepat dan cara kolaborasi antara metode satu dengan metode lainnya untuk meningkatkan kinerja peringkasan lebih baik.

2.1.2 *Preprocessing*

Preprocessing adalah tahapan awal mempersiapkan data. Data tidak terstruktur diubah menjadi data terstruktur sesuai dengan kebutuhan peringkasan.

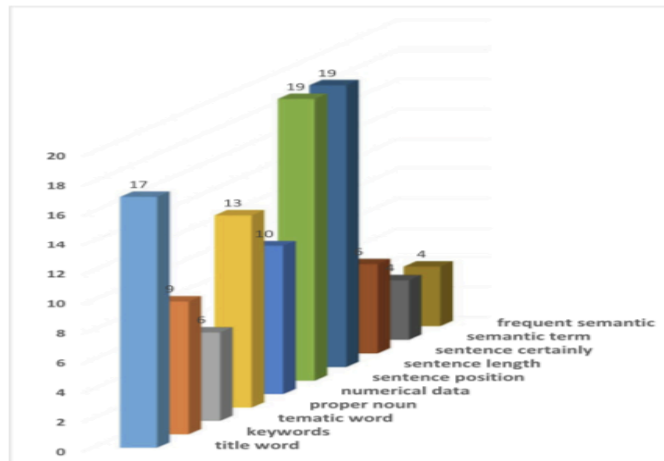


Gambar 3. Tahap *Preprocessing* Pada *Text Summarization*[12]

Berdasarkan kajian 10 tahun terakhir dapat disimpulkan bahwa *preprocessing* pada *text summarization* yang paling sering dilakukan adalah *stop word removal*. Stop word removal adalah penghapusan kata yang diabaikan saat pemrosesan dimana kata – kata yang diabaikan ini disimpan di daftar stop word. Ciri utama untuk menentukan stop word adalah kata – kata yang biasanya memiliki frekuensi kemunculan yang sering, misalnya konjungsi seperti “and”, “or”, “but”, “will” dan lain – lain. Tidak ada aturan pasti dalam menentukan stop word yang digunakan, penentuan stop word disesuaikan dengan kasus yang sedang diselesaikan dan sesuai dengan bahasa yang digunakan misalnya daftar stop word bahasa India akan berbeda dengan bahasa Inggris atau bahasa lainnya.

2.1.3 *Features*

Fitur dalam peringkasan teks adalah ciri atau indikator khusus yang akan diekstraksi untuk menghasilkan ringkasan. Terdapat 10 fitur yang sering digunakan untuk penelitian peringkasan teks dalam 10 tahun terakhir.

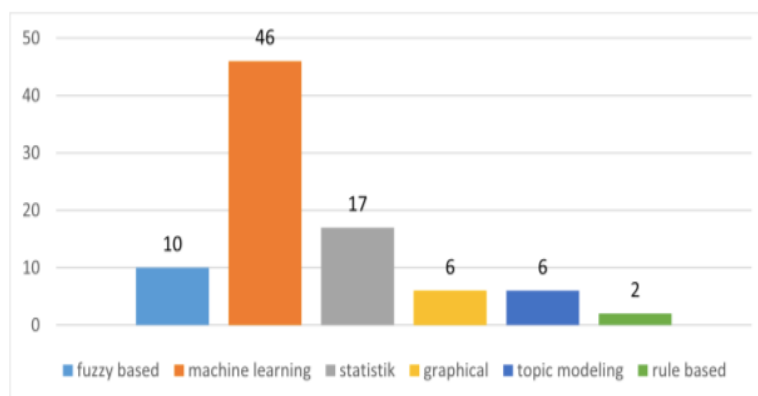


Gambar 4. *Features Text Summarization*[12]

Fitur yang paling sering digunakan adalah *sentence length* dan *sentence position*. Pada fitur *sentence length*, kalimat panjang mengandung informasi yang penting atau relevan yang artinya kalimat pendek tidak mencakup informasi yang relevan sehingga kalimat pendek dianggap tidak penting atau diabaikan. Beberapa metode telah digunakan untuk fitur *sentence position*, salah satunya yaitu melakukan pembobotan dengan urutan kalimat terbalik.

2.1.4 Teknik *Text Summarization*

Hasil literatur dari 10 tahun terakhir terdapat 6 pendekatan atau teknik peringkasan teks yang digunakan, yaitu *fuzzy based*, *machine learning*, *statistik*, *grafik*, *topic modeling* dan *rule based*.



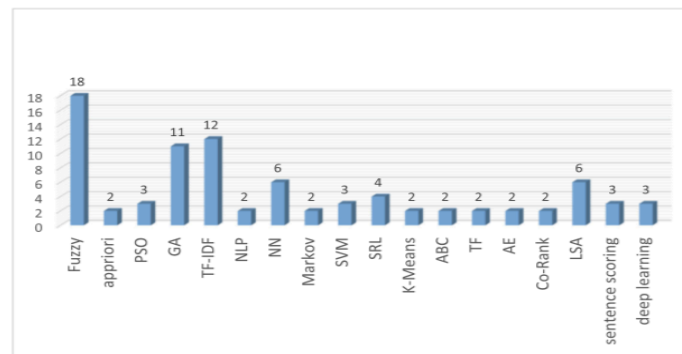
Gambar 5. Teknik *Text Summarization*[12]

Teknik yang paling sering digunakan pada peringkasan teks adalah *machine learning*. *Machine learning* merupakan teknik favorit karena merupakan teknik modern. Performa *machine learning* bersifat otomatis dan belajar improvisasi dari pengalaman tanpa diprogram secara eksplisit. Meski demikian *machine learning* bukanlah satu – satunya teknik terbaik.

Banyak penelitian dengan teknik machine learning memiliki kelemahan dalam meringkas teks seperti kelemahan dari segi semantik, memiliki kelemahan pada pengulangan kalimat dalam yang ringkasan yang artinya terkait dengan masalah redudansi kalimat.

2.1.5 Metode *Text Summarization*

Terdapat beragam metode peringkasan teks yang dapat digunakan. Berdasarkan studi literatur, metode yang paling banyak digunakan selama 10 tahun terakhir dalam peringkasan teks adalah fuzzy logic.

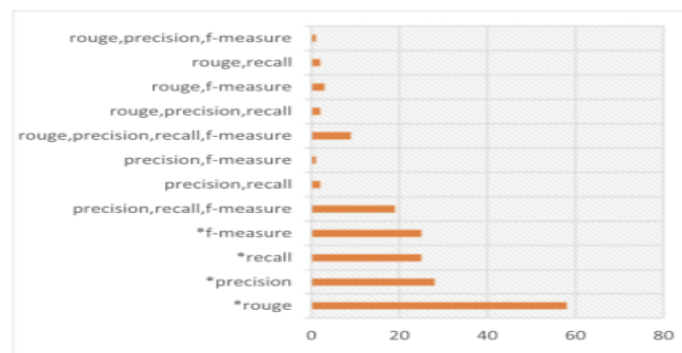


Gambar 6. Metode *Text Summarization*[12]

Metode fuzzy logic menjadi favorit yang digunakan karena sering digunakan atau mengekstrak dan menentukan nilai akhir dari kata atau kalimat yang termasuk dalam ringkasan. Metode fuzzy logic dapat mencegah kontradiksi data karena melibatkan peran manusia untuk dapat mengkaji kalimat dan mencapai kesepakatan atas pilihan kalimat tertentu untuk menghasilkan kalimat ringkasan.

2.1.6 Evaluasi Pada *Text Summarization*

Evaluasi adalah proses identifikasi untuk mengukur atau menilai kinerja suatu metode.



Gambar 7. *Evaluation Text Summarization*[12]

Dari studi literature 10 tahun terakhir, pendekatan evaluasi yang paling banyak dilakukan adalah dari segi ekstrak kalimat dan *content based*. Dalam hal ekstrak kalimat pengukuran yang sering dilakukan adalah *precision*, *recall*, *f-measure* / *f-score*. Sedangkan dari *content based* yang sering dilakukan adalah *N-Gram matching (Rogue)*.

2.1.7 Masalah Terkait *Text Summarization*

Berdasarkan studi literatur penelitian peringkasan teks dari 10 tahun terakhir, beberapa masalah menjadi kendala dalam peringkasan teks yang coba diselesaikan. Masalah yang paling sering dilakukan adalah ekstraksi karena merupakan masalah yang cukup *challenging*. Ekstraksi dalam *text summarization* adalah pengambilan data dari suatu sumber data baik data terstruktur atau data tidak terstruktur untuk diolah sehingga menjadi ringkasan. Teknik yang sering digunakan untuk masalah ekstraksi adalah *machine learning* dan ini sebanding dengan teknik yang paling sering digunakan dalam peringkasan teks adalah *machine learning*.

Masalah lain terkait penelitian peringkasan teks yang menjadi kelemahan pada penelitian 10 tahun terakhir adalah semantik. Permasalahan semantik perlu mendapatkan perhatian agar kalimat atau kata – kata yang termasuk dalam ringkasan sesuai dengan maksud dari dokumen tersebut. Beberapa peneliti menulis dalam penelitiannya yang dalam kesimpulan atau *future work* bahwa masalah yang belum diselesaikan dengan baik oleh para peneliti adalah semantik.

2.2 Perkembangan Penelitian Terkait

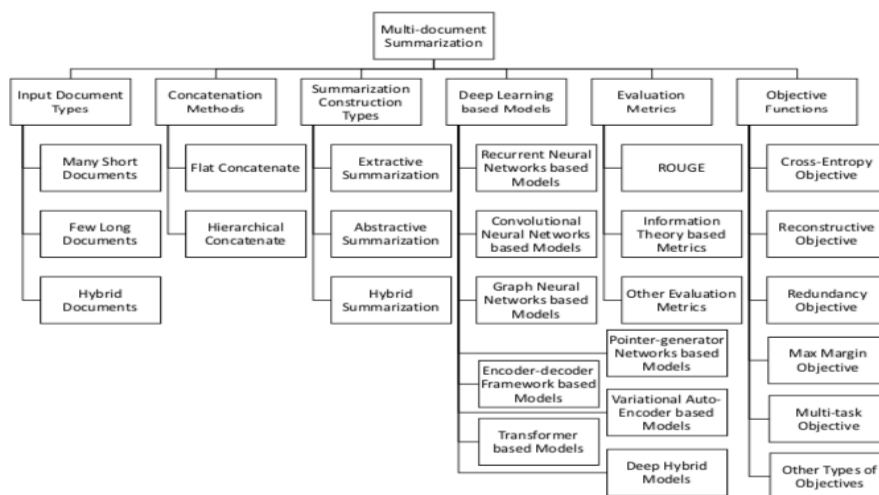
Peneliti yuliska, dkk [13] melakukan penelitian studi literatur mengenai metode, dataset, aplikasi dan teknik evaluasi yang dapat diimplementasikan untuk riset dibidang peringkasan teks dokumen berbahasa Inonesia. Hasil studi literatur menemukan bahwa peringkasan teks secara otomatis didominasi oleh teknik ekstraktif. Metode peringkasan teks berbahasa Indonesia paling sering digunakan adalah metode *unsupervised*, sedangkan metode *supervised* masih sangat jarang ditemukan. Parameter evaluasi yang lebih tepat digunakan untuk mengevaluasi hasil ringkasan teks adalah ROGUE.

Peneliti Rike Adelia, dkk [14] melakukan penelitian peringkasan dokumen teks bahasa Indonesia secara abstraktif dengan metode RNN. Peneliti menyatakan bahwa penelitian peringkasan teks secara abstraktif memiliki kesulitan lebih daripada secara ekstraktif karena menyusun ringkasan dari seluruh isi teks memiliki kesulitan yang tinggi. Peneliti melakukan penelitian terhadap 1000 dokumen artikel ilmiah bahasa Indonesia. Hasil penelitian menunjukkan bahwa model dapat mempelajari dan memahami kata – kata pada sumber

dokumen serta menghasilkan ringkasan. Akan tetapi hasil ringkasannya memiliki struktur kalimat kurang baik. Dari hasil penelitian diketahui bahwa nilai evaluasi metode RNN dalam peringkasan teks dokumen dalam bahasa Indonesia tidak lebih tinggi dari peringkasan teks dokumen bahasa Inggris dan bahasa China.

Peneliti Khairiyyah Nur Aisyah, dkk [15] melakukan penelitian peringkasan multi-dokumen berdasarkan metode *sentence extraction* dan *word sense disambiguation* dengan teknik clustering sentence. Hasil ringkasan dari 50 topik berita kemudian dihitung menggunakan nilai ROGUE. Dari hasil pengujian didapat bahwa peringkasan teks multi-dokumen hanya dengan menggunakan metode *sentence extraction* memiliki nilai ROGUE-1 sebesar 0.68 dan nilai ROGUE-2 0.32. Sedangkan peringkasan multi-dokumen dengan menggunakan metode *sentence extraction* dan metode *word sense disambiguation* memiliki nilai ROGUE-1 sebesar 0.71 dan nilai ROGUE-2 sebesar 0.36. Peringkasan teks multi-dokumen dengan memperhatikan *sentence extraction* dan *word sense disambiguation* memiliki hasil nilai ROGUE-1 dan ROGUE-2 yang lebih baik meski memiliki nilai perbandingan yang tidak terlalu tinggi. Hal ini bias jadi dikarenakan kata polisemi yang dimiliki set dokumen tidak terlalu banyak sehingga hasil ROGUE-1 dan ROGUE-2 nya tidak memiliki perbedaan yang cukup signifikan. Untuk penelitian selanjutnya diharapkan dapat melakukan peringkasan teks multi-dokumen dengan memperhatikan *word sense disambiguation* dan sinonim kata.

Peneliti Congbo MA, dkk [9] melakukan penelitian survei *multi-document summarization* dengan teknik *deep learning*. Hasil survey menyatakan bahwa model *transformer based* adalah model *multi-document summarization* yang paling sering digunakan selama dua tahun terakhir. *Graph neural networks* (GNN) merupakan arsitektur yang ideal untuk pemrosesan natural language karena mampu memodelkan keterhubungan yang kuat antara entitas secara semantik dan sintaksis. *Cross-entropy* adalah *objective function* yang paling sering digunakan. Sedangkan parameter evaluasi ROGUE series adalah salah satu parameter *multi-document summarization* yang secara umum diterima oleh komunitas peneliti di bidang *natural language processing*. Akan tetapi parameter lain seperti BLEU dan METEOR memiliki potensi untuk digunakan sebagai parameter evaluasi dalam *multi-document summarization*.



Gambar 8. Struktur Hirarki Survey[9]

Penelitian – penelitian terkait peringkasan teks multi-dokumen atau multi-document summarization disajikan pada tabel 2.1.

Tabel 2.1 Tabel Perbandingan Penelitian

Nama / Judul	Tujuan / Metode	Hasil	Peluang pengembangan
Yuliska, Khairul Umam Syaliman Literatur review terhadap metode, aplikasi dan dataset peringkasan dokumen teks otomatis untuk teks berbahasa Indonsia.(2020)	Melakukan penelitian studi literatur mengenai metode, dataset, aplikasi dan teknik evaluasi yang dapat diimplementasikan untuk riset dibidang peringkasan teks dokumen berbahasa Inodonesia.	Pemrosesan bahasa Indonesia dan bahasa Inggris berbeda terutama pada proses <i>preprocessing</i> . Teknik dan metode peingkasan teks berbahasa Indonesia terpaku pada metode – metode perangkingan tradisional seperti	Pengembangan metode peringkasan teks berbahasa Indonesia untuk metode <i>machine learning</i> atau <i>deep learning</i> .

		<p><i>maximal marginal relevance</i> (MMR) dan <i>text rank</i>.</p> <p>Untuk evaluasi <i>measure</i>, riset peringkasan teks berbahasa Indonesia lebih banyak menggunakan <i>recall</i> dan <i>precision</i>.</p> <p>Evaluasi <i>measure</i> yang tepat untuk peingkasan teks adalah ROGUE.</p>	
<p>Rieke Adelia, Suyanto, Untari Novia Wisesty</p> <p>Indonesian Abstractive Summarization Using Bidirectional Gated Recurrent Unit. (2019)</p>	<p>Melakukan peringkasan teks artikel ilmiah bahasa Indonesia</p> <p>- <i>Abstractive Summarization</i></p> <p>- RNN</p>	<p>Model dapat mempelajari dan memahami kata – kata serta menghasilkan hasil ringkasan. Akan tetapi struktur kalimat hasil ringkasannya kurang baik.</p> <p>Hasil score evaluasi jauh di bawah dari score</p>	<p>Untuk penelitian selanjutnya agar mengatasi masalah linguistik dan menangani struktur tata bahasa untuk menghasilkan hasil ringkasan yang lebih baik.</p>

		evaluasi ringkasan abstraktif bahasa Inggris dan China dengan metode RNN.	
Khairiyyah Nur Aisyah, Syadza Anggraini, Agus Zainal Arifin Peringkasan Teks Multi-Dokumen Dengan Metode Sentence Extraction dan Word Sense Disambiguation. (2019)	Mengusulkan metode peringkasan teks multi-dokumen yang dapat meningkatkan relevansi antara kalimat dengan : - Metode <i>Sentence Extraction</i> - Metode <i>Word Sense Disambiguation</i> - Metode <i>Clustering</i>	Pringkasan teks multi-dokumen dengan metode <i>sentence extraction</i> dan metode <i>word sense disambiguation</i> berhasil dilakukan dan memiliki Nilai ROGUE-1 sebesar 0.71 dan ROGUE-2 sebesar 0.36. Artinya peringkasan teks multi-dokumen dengan memperhatikan juga <i>word sense disambiguation</i> menghasilkan ringkasan yang lebih baik.	Melakukan pengembangan penelitian peringkasan teks multi-dokumen dengan memperhatikan <i>word sense disambiguation</i> dan sinonim kata.
Congbo MA, Wei Emma Zhang, Mingyu Guo, Hu Wang, Quan Z. Sheng	Melakukan studi literatur terkait peringkasan <i>multi-document</i> secara ekstraktif dengan	<i>Transformer based</i> model adalah model yang paling banyak digunakan	<i>Multi-document summarization</i> masih memiliki peluang untuk dijadikan topik penelitian karena topik

Multi-document Summarization Via Deep Learning Techniques: A Survey. (2020)	teknik <i>Deep Learning</i> .	<p>hampir selama 2 tahun.</p> <p><i>Graph neural networks</i> (GNN) merupakan arsitektur yang ideal untuk pemrosesan <i>natural language</i> karena mampu memodelkan keterhubungan yang kuat antara entitas secara semantik dan sintaksis.</p> <p>Parameter evaluasi <i>ROGUE series</i></p>	ini masih jarang dibandingkan dengan <i>single documentation</i> .
---	-------------------------------	--	--

Bab 3

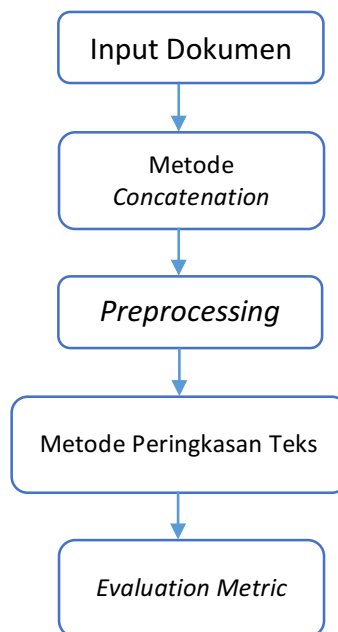
Metodologi

3.1 Objek Penelitian

Sesuai dengan tema yang diusulkan yaitu peringkasan teks multi-dokumen untuk pengembangan *smart content* di dunia Pendidikan maka objek dari penelitian ini adalah dokumen teks *ebook* berbahasa Indonesia yang didapat dari internet.

3.2 Tahapan Penelitian

Penelitian ini berusaha mengembangkan metode peringkasan teks multi-dokumen dengan teknik ekstraktif untuk mendapatkan ringkasan teks yang lebih baik. Tahapan yang akan dilakukan pada penelitian seperti tampak pada gambar 9.



Gambar 9. *Framework Proses Multi-Document Summarization*

Tahapan – tahapan penelitian yang akan dilakukan adalah input dokumen, metode *concatenation*, *preprocessing*, metode peringkasan teks dan tahap akhir yaitu evaluasi hasil ringkasan.

3.2.1 Input Dokumen

Pada peringkasan teks multi-dokumen tipe dan panjang dari input dokumen beragam dan dibagi menjadi 3 bagian yaitu, *many short document*, *few long document* dan *hybrid document*. Ringkasan teks yang dibentuk pada penelitian ini adalah ringkasan ilmiah dari dokumen yang panjang dengan objek penelitian ebook atau tulisan ilmiah maka tipe input dokumennya adalah *hybrid document*[16].

3.2.2 Metode Concatenation

Jumlah input dokumen yang besar mungkin mengandung kontradiksi, redudansi, dan informasi yang saling melengkapi. Untuk menangani hubungan yang kompleks pada input multi dokumen diperlukan metode concatenation untuk mengetahui hubungan antara dokumen. Terdapat dua metode concatenation untuk menggabungkan beberapa dokumen input yaitu *flat concatenation* dan *hierarchical concatenation*. Pada penelitian ini metode concatenation yang rencananya akan digunakan adalah *hierarchical concatenation* dengan metode *graph based* karena metode ini mampu mempertahankan *cross document relations*. Hubungan hirarki antara dokumen dapat memfasilitasi model untuk mendapatkan representasi semantik yang baik yang dapat meningkatkan efektivitas model [11].

3.2.3 Preprocessing

Preprocessing adalah proses awal untuk mempersiapkan data. Data yang tidak terstruktur dirubah menjadi data terstruktur disesuaikan dengan kebutuhan proses peringkasan. Pada penelitian ini tahapan preprocessing yang akan dilakukan adalah *stop word removal*, *stemming*, *tokenizing*, dan *segmentation* [7].

3.2.4 Metode Peringkasan Teks

Metode peringkasan yang diusulkan pada penelitian ini adalah *graph neural network* (GNN).

3.2.5 Metric Evaluasi

Metric evaluasi adalah guna mengukur efektivitas dari metode dengan cara yang objektif. Metric evaluasi yang digunakan pada rencana penelitian ini adalah ROGUE series karena ROGUE series merupakan parameter evaluasi yang tepat digunakan dalam peringkasan teks dokumen.

3.3 Rencana Kerja

Rencana kerja adalah perencanaan kegiatan yang akan dilakukan pada penelitian peringkasan teks multi-dokumen. Berikut ini adalah uraian rencana kerja yang disusun sesuai dengan kegiatan yang akan dilakukan.

Tabel 3.1 Rencana Kerja Penelitian

Kegiatan	Bulan Ke-																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Studi Pustaka																						
Pemilihan Topik																						
Eksplorasi Objek																						
Penyusunan Proposal																						
Pengajuan Proposal																						
Pengerjaan Penelitian																						
Publikasi Ilmiah/Seminar																						
Penyusunan Disertasi																						
Sidang Disertasi																						

Daftar Pustaka

- [1] I.E. Bisen, F. Nalcaci, S. Alagappan, and Y. Yildirim, "The state of Education and artificial Intelligence after the pandemic," in *London Journal of Social Sciences*, vol.1, pp. 88-98,2021.
- [2] W. Holmes, M. Bialik, C. Fadel, "Artificial intelligence in Education," in *Boston: Center for Curriculum Redesign*, 2019.
- [3] R. Bajaj, V. Sharma, "Smart Education with artificial intelligence based determination of learning styles," in *International Conference on Computational Intelligence and Data Science*, vol. 132, pp. 834-842, 2018.
- [4] M. Fahimirad, S.S. Kotamjanji, "A review on application of artificial intelligence in teaching and learning in Educational Context," in *International Journal of Learning and Development*, vol.8, pp. 106-118, 2018.
- [5] S. Fatimah, W. Setiawan, E. Junaeti, and A.S. Surur, "Development of smart content model-based augmented reality to support smart learning," in *Journal of Science Learning*, vol. 2, pp. 65-70, 2019.
- [6] N. Lin, J. Li, S. Jiang, "A simple effective method for Indonesian automatic text summarization," in *Connection Science*, 2021.
- [7] R. Xu, J. Cao, M. Wang, J. Chen, H. Zhou, Y. Zheng, Y. Wang, L. Chen, X. Yin, X. Zhang, S. Jiang, Y. Wang, and L. Li, "Xiaomingbot: a multilingual robot news reporter," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020.
- [8] G. Garmaset, M.L. Khodra, "Summarizing Indonesian news articles using graph convolution network," in *Journal of Information and Communication Technology*, vol.18, pp. 345-365
- [9] M.A. Congbo, W.E. Zhang, M. Guo, H. Wang, and Q.Z. Sheng, "Multi-document summarization via deep learning techniques: a survey," in *Assosiation for Computing Machinery*, vol.1, pp. 1-35, 2020.
- [10] R. Nallapati, F. Zhai, B. Zhou, "SummaRuNNer: A recurrent neural network based sequence model for extractive summarization of documents," in *Proceedings of the Thirty First AAAI Conference on Artificial Intelligence*, 2017.

- [11] R. Paulus, C. Xiong, R. Socher, "A deep reinforced model for abstractive summarization," in proceedings of the 6th International Conference on Learning Representation, 2018.
- [12] A.P. Widyassari, S. Rustad, G.F. Shidik, E. Noersasongko, A. Syukur, A. Afandy, and D.R.I.M. Setiadi, "Review of automatic text summarization techniques and methods," in Journal of King Saud University-Computer and Information Sciences, 2020.
- [13] Yuliska, K.U. Syaliman, "Literatur review terhadap metode, aplikasi dan dataset peringkasan dokumen teks otomatis untuk teks Berbahasa Indonesia," in IT Journal Research and Development (ITJRD), vol. 5, pp. 19-31, 2020.
- [14] R. Adelia, S. Suyanto, U.N. Wisesty, "Indonesian abstractive text summarization using bidirectional gated recurrent unit," in procedia Computer Science, vol. 157, pp. 581-588, 2019.
- [15] K.N. Aisyah, S. Anggraini, A.Z. Arifin, "Peringkasan teks multi-dokumen berdasarkan metode sentence extraction dan word sense disambiguation," in NJCA, vol. 4, pp. 1-8, 2019.