

SEMANTIC RULE LABELING DAN SENTENCE INFORMATION DENSITY DALAM PEMILIHAN KALIMAT REPRESENTATIF CLUSTER PADA PERINGKASAN MULTI-DOKUMEN

Gus Nanang Syaifuddiin¹, Agus Zainal Arifin², Diana Purwitasari³
Jurusan Teknik Informatika, Fakultas Teknologi Informasi, Institut Teknologi Sepuluh Nopember
Email: 9uz.4nang@gmail.com¹, agusza@cs.its.ac.id², diana@if.its.ac.id³

ABSTRAK

Coverage dan *salient* merupakan masalah utama yang menjadi perhatian para peneliti dalam peringkasan dokumen. Pendekatan *clustering* mampu memberikan *coverage* yang baik terhadap semua topik namun tidak memiliki informasi-informasi yang bisa mewakili kalimat-kalimat lain (*salience sentence*). *Salience* dapat digali dengan melihat hubungan dari satu kalimat dengan kalimat lain yang dibangun dengan pendekatan *position text graph*, namun *position text graph* hanya mampu menggali hubungan antar kalimat tanpa memperhatikan peran semantik kata ("who" did "what" to "whom", "where", "when", and "how") dalam kalimat yang dibandingkan. Pada paper ini kami mengusulkan sebuah metode baru strategi pemilihan kalimat representatif *cluster* dengan pendekatan *sentence information density* dan *Semantic Rule labeling*. Hasil uji coba menunjukkan metode yang diusulkan mampu memilih kalimat ringkasan lebih baik dari metode *Sentence Information Density* (SID) dengan rata-rata nilai Rouge-1 0.32511.

Kata Kunci: Peringkasan Multi-Dokumen, *Position Text Graph*, *Semantic Role Labeling*, *Salience* Dan *Coverage*.

ABSTRACT

Coverage and *salient* is the main problem to the attention of researchers in peringkasan document. Clustering approach capable of providing good coverage of all the topics but does not have information that can represent other sentences (*salience sentence*) .*Salience* can be explored by looking at the relationship from one sentence to another sentence that was built with the approach of graph text position, but *position text graph* only be able to explore the relationship between a sentence without considering the role of semantic word ("who" did "what" to "whom", "where", "when" and "how") in the sentence being compared. In this paper we propose a new method of election strategy representative sentence *sentence cluster* approach *information density* and *Semantic Rule labeling*. The trial results showed that the method proposed method is able to choose better-sentence summary of the method *Sentence Information Density* (SID) with an average value of Rouge-1 0.32511.

Keywords: Multi-Document Summarization, *Position Text Graph*, *Semantic Role Labeling*, *Salience* and *Coverage*.

1. Pendahuluan

Terdapat dua metode yang digunakan dalam melakukan peringkasan multi-dokumen secara otomatis: *abstractive* dan *extractive*. Peringkasan secara *abstractive* dilakukan dengan mendapatkan informasi

yang disampaikan oleh dokumen sumber dan membentuk peringkasan dengan teknik menyatukan atau menurunkan informasi [1].

Peringkasan *extractive* dilakukan dengan melakukan ekstraksi terhadap

kalimat dan mengurutkan berdasarkan nilai paling tinggi dan dijadikan sebagai kandidat ringkasan. Sebagian besar peneliti terfokus pada peringkasan multi-dokumen secara extractive seperti yang dilakukan pada penelitian ini.

Kalimat yang dipilih pada hasil ringkasan harus mempunyai good coverage dan salient terhadap topik dari dokumen sumber. Coverage dan salient menjadi masalah utama dalam peringkasan secara abstractive atau extractive. Clustering kalimat merupakan salah satu metode yang dapat memberikan good coverage. Beberapa penelitian diantaranya dengan CLASSY (clustering, Linguistics, And Statistics for Summarization Yield) [2] digunakan untuk melakukan pemangkasan kalimat secara bahasa dan menggunakan metode statistik untuk mendapatkan ringkasan topik dari dokumen sumber.

Ma dkk [3] membangun cluster berdasarkan gabungan query sentence dan mendapatkan ringkasan kalimat berdasarkan modifikasi MMR (Maximal Marginal Relevance)[4]. Gupta dkk [5] menyajikan sebuah metode untuk peringkasan multi-dokumen dengan menggabungkan ringkasan dokumen tunggal dan membentuk cluster kalimat dari beberapa fitur diantaranya bobot, kalimat, lokasi fitur dan konsep kesamaan fitur.

Good coverage dapat dicapai dengan menjaga tingginya hubungan antar cluster [6], sehingga dapat mengidentifikasi topik dan sub-topik kedalam cluster-cluster bentukan. Koherensi dari cluster secara dinamis dipantau dengan menggunakan konsep yang disebut cluster similarity histogram [7]. Similarity based Histogram clustering (SHC) terbukti lebih baik jika dibandingkan dengan Hierarchical Agglomerative clustering (HAC), Single-Pass clustering dan K-Nearest Neighbor clustering .

Banyak peneliti mengidentifikasi salient sentence dengan meningkatkan metode sentence ranking. Hal ini juga diakui

bahwa strategi pemilihan kalimat sangat penting dan bertujuan untuk mengurangi redundansi antara kalimat yang dipilih sehingga hasil ringkasan memungkinkan dapat mencakup lebih banyak konsep dari dokumen sumber. Beberapa metode telah dikembangkan antara lain [8] mengurangi bobot dari kalimat yang mengandung discourse connectors (DC) seperti “because”, “as a result”, “after” dan “before”.

Kalimat penting (*salient sentence*) penyusun sebuah ringkasan harus memiliki kepadatan informasi. *Salient sentence* harus mengandung informasi sebanyak mungkin dari dokumen sumber (He dkk, 2008). Menurut He dkk (2008) fitur kepadatan informasi kalimat sentence information density (SID) dapat digali dengan pendekatan positional text graph. Pendekatan positional text graph digabungkan dengan kata kunci cluster untuk meningkatkan pembobotan kalimat dalam *cluster* [16].

Dibidang Natural Language Processing (NLP) telah dikembangkan sebuah metode semantik parsing yang diberi nama *Semantic Role Labeling* (SRL) [10] yang dapat digunakan untuk mengidentifikasi argumen dari predikat dalam suatu kalimat, dan menentukan semantic role atau peran semantik.

Sebagian besar algoritma penyusunan peringkat berbasis graph menggunakan kalimat sebagai bag of word atau informasi sintaksis dalam dokumen teks, tapi mengabaikan informasi semantik. Manusia memahami kalimat berdasarkan peran semantik kata (“who” did “what” to “whom”, “where”, “when”, and “how”). Informasi semantik lebih sesuai untuk menggambarkan persepsi manusia terhadap kalimat dibandingkan dengan kalimat sebagai *bag of word*.

Oleh karena itu pada peneitian ini kami mengusulkan metode baru, strategi pemilihan kalimat representatif *cluster* dengan pendekatan position text graph dan

peran semantik kata pada peringkasan multi-dokumen. Sehingga dapat meningkatkan kemungkinan kesesuaian hasil ringkasan dan menghilangkan kalimat yang tidak relevan pada hasil ringkasan.

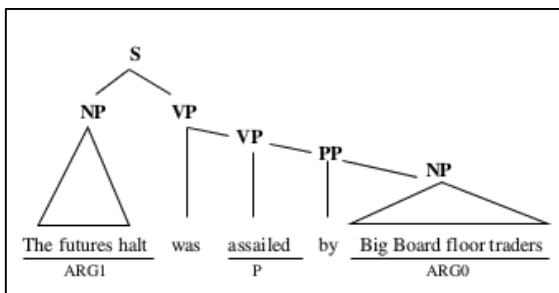
3. Semantic Role Labelling (SRL)

SRL adalah suatu proses yang digunakan untuk menentukan hubungan peran semantik antar kata dalam suatu kalimat. Baker [11] mengemukakan gagasan *Framenet* untuk menggambarkan fungsi tata bahasa, Jenis frase, dan sifat-sifat sintaksis lainnya. *Framenet* difokuskan pada *frame semantic* yang merepresentasikan skematis dari situasi yang digambarkan suatu kalimat.

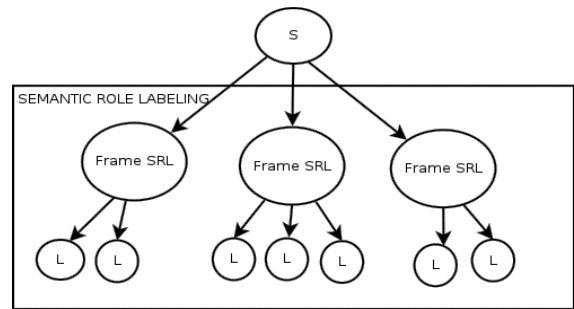
Bjorkelund [12] memberikan kontribusi dengan menambahkan klasifikasi untuk mengidentifikasi arti dari predikat, arti dari argument dan label dari argument yang kemudian didemonstrasikan dalam penelitian [13].

Pada penelitian lain Palmer dkk [14] mendiskusikan peran semantik yang digunakan untuk proses anotasi yang disebut *proposition bank (Probank)* dengan pendekatan penyediaan data statistik untuk memberikan penjelasan pada setiap klausa *treeBank*.

Extraksi *Semantic Role Labeling* dapat dibagi menjadi dua sub tasks utama, yaitu mengidentifikasi *Frame SRL* dalam kalimat dan melabeli kata pada *Frame SRL* dengan label yang semantik yang telah disediakan SRL.



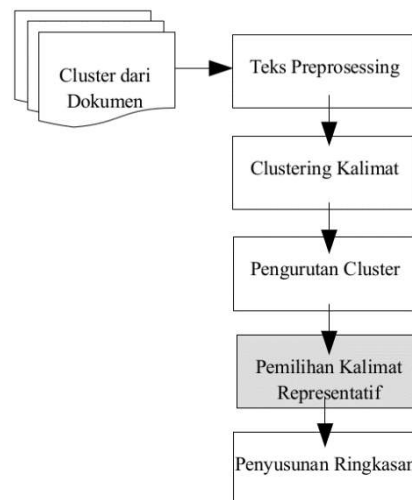
Gambar 1. Proses *Semantic Role Labeling*



Gambar 2 Extraksi Kalimat dengan *Semantic Role Labeling*

4. Model Sistem

Desain model sistem yang digunakan diadopsi dari *framework* [6] yang ditunjukkan pada Gambar 1.



Gambar 3. Desain Model Sistem[6]

Pertama data yang telah berupa *cluster* dari dokumen dilakukan *text preprocessing* untuk memudahkan dalam pengolahan data selain itu juga mengekstraksi kata dan kalimat yang berada pada tiap dokumen, tahapan selanjutnya adalah data dari hasil preprocessing dibentuk menjadi *cluster-cluster* kalimat dengan menggunakan *Similarity Histogram Cluster (SHC)*.

3.1 Teks Preprocessing

Proses yang dilalui dalam *teks preprocessing* adalah *segmentation*, *stopword removal*, dan *stemming*. Pada preprocessing ini dilakukan ekstraksi

Semantic Rule labeling dari setiap kalimat yang kemudian disimpan dan akan digunakan pada pemilihan kalimat representative.

3.2 Clustering Kalimat

Data dari hasil preprocessing selanjutnya dilakukan *clustering* berdasarkan *Similarity Histogram Cluster* (SHC) [6]. *Similarity Histogram Cluster* merupakan representasi statistik dari suatu distribusi similaritas pasangan antar anggota yang ada pada suatu *cluster*.

Jumlah dari *bin* dalam histogram menunjukkan interval nilai similaritas tertentu. Namun *vector* yang menunjukkan similaritas dari kalimat-kalimat menjadi sangat jarang (*too sparse*) karena kalimat terlalu pendek ketika dibandingkan satu dengan yang lain. Sehingga Sarkar [6] menggunakan *uni-gram matching-based similarity measure* ditunjukkan pada Persamaan 1:

$$\text{sim}(s_i, s_j) = \frac{(2 * |s_i| \cap |s_j|)}{|s_i| + |s_j|} \quad (1)$$

dimana s_i dan s_j adalah kalimat s ke- i dan ke- j . Selanjutnya $|s_i| \cap |s_j|$ merepresentasikan jumlah dari kata-kata yang sesuai antara kalimat s ke- i dan kalimat s ke- j . $|s_i|$ adalah panjang kalimat s ke- i yaitu jumlah kata yang menyusun kalimat tersebut.

Kualitas hubungan antar *cluster* dinilai dengan menghitung rasio jumlah kesamaan diatas *similarity threshold* S_T terhadap total kesamaan. Jika n_c adalah jumlah dari kalimat pada suatu *cluster*, maka jumlah dari pasangan kalimat yang ada pada *cluster* tersebut adalah $m_c = n_c(n_c + 1)/2$, dimana $S = \{s_i; i=1, \dots, m_c\}$ adalah himpunan kesamaan pada *cluster*. *Similarity histogram* dari *cluster* dinotasikan dengan $H = \{h_1, h_2, h_3, \dots, h_{n_b}\}$. Jumlah dari *bin* yang ada pada suatu histogram dinotasikan dengan n_b sedangkan jumlah similaritas kalimat yang ada pada *bin* ke- i dinotasikan

dengan h_i . Jumlah *bin* yang digunakan dalam penelitian ini adalah 10. Fungsi untuk menghitung h_i ditunjukkan pada Persamaan (2.2).

$$h_i = \text{count}(\text{sim}_j) \quad (2)$$

Untuk $\text{sim}_{l_i} \leq \text{sim}_j \leq \text{sim}_{u_i}$,

Sim_{l_i} adalah batas bawah similarity pada *bin* ke- i sedangkan sim_{u_i} adalah batas atas similaritas pada *bin* ke- i . *Histogram Ratio* (HR) dari suatu *cluster* dapat dihitung dengan Persamaan (2.3).

$$\text{HR} = \frac{\sum_{i=T}^{n_b} h_i}{\sum_{j=1}^{n_b} h_j} \quad (3)$$

$$T = \lfloor S_T * n_b \rfloor \quad (4)$$

S_T adalah *similarity threshold*. Persamaan (4) menunjukkan jumlah *bin* yang sesuai dengan *similarity threshold* yang dinotasikan dengan T .

3.3 Pengurutan Cluster

Untuk menatasi masalah ini maka diperlukan pengurutan *cluster* berdasarkan *cluster important* [6]. Jika frekuensi suatu kata $w(\text{count}(w))$ memenuhi *threshold* θ maka kata tersebut adalah kata *frequent*. Dengan cara mengukur kata bobot *cluster* berdasarkan kata-kata penting yang terdapat didalamnya. Pentingnya *cluster* dihitung dengan Persamaan 5.

$$W(c_j) = \sum_{w \in c_j} \log(1 + \text{count}(w)) \quad (5)$$

Dimana bobot *cluster* c ke- j dinotasikan dengan $W(c_j)$ dan $\text{count}(w)$ adalah jumlah kata pada koleksi *input* dan $\text{count}(w)$ lebih dari *threshold* θ . Pentingnya kata diukur dengan nilai *log* yang menormalisasikan jumlah total kata pada kumpulan seluruh dokumen *input* setelah dilakukan proses *stopwords removal*.

Setelah melakukan pengurutan *cluster* secara *descending*, *top cluster* dipilih

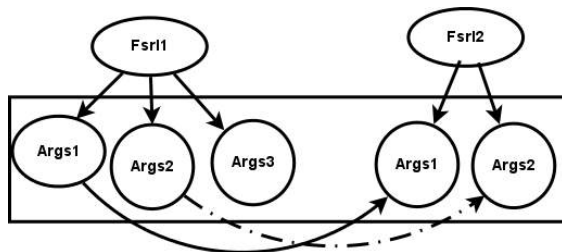
sebagai kandidat topik yang mewakili topik-topik dari dokumen *input* dan selanjutnya dilakukan pemilihan kalimat representatif yang bisa mewakili *top cluster*.

3.4 Pemilihan Kalimat Representatif

Pada pemilihan kalimat representatif similaritas yang digunakan adalah kombinasi hasil ekstraksi dari SRL dan metode *Position text graph* yang dibangun berdasarkan konsep similaritas antar setiap kalimat yang berada di dalam *cluster*.

Setiap kalimat yang ada dianalogikan sebagai sebuah *vertex* dan hubungan *similarity* antar kalimat dinyatakan dengan *edge* yang menghubungkan kalimat-kalimat tersebut. *Edge* antar kalimat terbentuk jika similaritas antara dua kalimat lebih besar dari *threshold* (α) yang ditentukan.

Pada pengukuran jarak similaritas antara kalimat digunakan pengukuran *Jaccard coefficient* Persamaan (6).

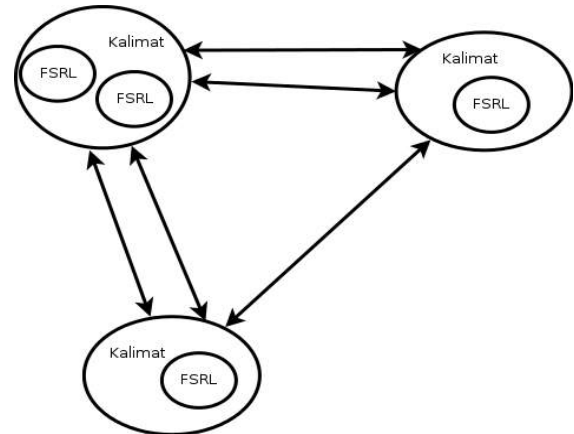


Gambar 4. Similaritas Label dari Hasil *Semantic Rule Labeling*

Hasil ekstraksi SRL yang berupa *Frame SRL* dan peran semantik dijadikan fitur yang akan digunakan dalam penghitungan *position text graph*. Pada perhitungan awal perbandingan peran semantik dimana tiap argumen dibandingkan dengan kata atau kalimat yang mempunyai argumen yang sama. Pada perbandingan *Args1* sampai *Args7* digunakan *Jaccard coefficient* dengan sedikit modifikasi dan *threshold*.

$$SimArgs(Args_{si}, Args_{di}) = \frac{C(Args_{si}) \cap C(Args_{di})}{C(Args_{si}) \cup C(Args_{di})} \quad (6)$$

Dimana $Args_{si}$ adalah argumen dari hasil ekstraksi *srl* ke-s pada argumen ke-i jika $SimArgs(Args_{si}, Args_{di})$ lebih besar dari *threshold* T maka $SimArgs(Args_{si}, Args_{di})=1$ jika tidak $SimArgs(Args_{si}, Args_{di})=0$.



Gambar 5. *Sentence Information Density* yang diusulkan

Sehingga didapat jarak antar kalimat adalah penggabungan jarak antar frame dalam kalimat tersebut.

3.5 Penyusunan Ringkasan

Sebuah kalimat representatif dipilih dari setiap *cluster* berdasarkan hasil dari proses pemilihan kalimat representatif. Pemilihan kalimat dimulai dari *cluster* yang memiliki bobot *cluster importance* paling tinggi. Kemudian pemilihan dilanjutkan pada *cluster* berikutnya sesuai dengan daftar urutan *cluster* berdasarkan bobot *cluster importance* secara *descending*.

4 Evaluasi Hasil

Kualitas hasil ringkasan pada penelitian ini dievaluasi dengan ROUGE. Metode ROUGE telah diadopsi dari DUC untuk mengevaluasi peringkasan teks otomatis. ROUGE sangat efektif digunakan untuk mengevaluasi peringkasan dokumen [15].

ROUGE mengukur kualitas hasil ringkasan dengan menghitung unit-unit yang *overlap* seperti *N-gram*, urutan kata dan pasangan-pasangan kata antara ringkasan kandidat dan ringkasan sebagai referensi. ROUGE- N mengukur

perbandingan *N-gram* dari ringkasan, dan menghitung berapa jumlah yang cocok. Perhitungan ROUGE-*N* yang diadopsi dari perhitungan [15] ditunjukkan pada Persamaan (7):

$$ROUGE - N = \frac{\sum_{S \in \text{Summ}_{ref}} \sum_{N\text{-gram} \in S} \text{Count}_{match}(N\text{-gram})}{\sum_{S \in \text{Summ}_{ref}} \sum_{N\text{-gram} \in S} \text{Count}(N\text{-gram})} \quad (7)$$

Dimana *N* menunjukkan panjang dari *N-gram*, $\text{Count}_{match}(N\text{-gram})$ adalah jumlah maksimum dari *N-gram* yang muncul pada ringkasan kandidat dan ringkasan sebagai referensi. $\text{Count}(N\text{-gram})$ adalah jumlah dari *N-gram* pada ringkasan sebagai referensi.

Pada penelitian ini fungsi ROUGE yang digunakan adalah ROUGE dengan nilai *N-gram* =1. *N-gram* =1 dipilih karena ROUGE-1 lebih berkorelasi dengan ringkasan secara manual [15]. Hasil dari perhitungan ROUGE mempunyai nilai

Notasi	Inisialisasi
HR_{min}	0.4,0.5,0.6
ϵ	0.6,0.5,0.4
S_T	0.3,0.4,0.5,0.6
θ	10,12,15,17,20,22,25,27,30
α	0.4,0.5,0.6,0.7,0.8
T	0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9

kesamaan 0 sampai 1. Metode yang diusulkan akan dibandingkan dengan pendekatan positional text graph [16].

5. Uji Coba dan Hasil

Uji coba sistem dilakukan untuk menguji atau menjalankan sistem dengan beberapa parameter yang ada pada metode. Pada tahap uji coba enam buah parameter (HR_{min} , ϵ , S_T , θ , α dan T) sistem terlebih dahulu diestimasi melalui proses estimasi parameter.

Tujuan dari proses estimasi parameter parameter tersebut adalah mendapatkan nilai parameter-parameter yang paling optimal sehingga dapat memberikan hasil testing yang terbaik. Parameter-parameter yang terdapat pada sistem peringkasan ditunjukkan pada Tabel 1.

Alur pengujian sistem mulai dari estimasi parameter hingga testing sistem. Sebelum melakukan uji coba. Pada penelitian ini kinerja dari metode yang diajukan dievaluasi berdasarkan nilai ROUGE-N [15].

5.1 Estimasi Parameter Optimal

Proses estimasi parameter bertujuan untuk mencari nilai optimal dari parameter HR_{min} , ϵ , S_T , θ , dan α pada metode yang diusulkan. Pada proses estimasi parameter nilai parameter akan dilakukan inisialisasi dengan beberapa parameter.

Parameter yang telah diinisialisasi dikombinasikan untuk mendapatkan kombinasi nilai parameter terbaik yang akan digunakan untuk proses testing. Nilai inisialisasi parameter HR_{min} , ϵ , S_T , θ , dan α pada metode yang diusulkan juga akan digunakan pada pada metode [16] lain yang akan dibandingkan.

Hasil kombinasi nilai parameter yang terbaik dari tiap metode baik metode yang diusulkan atau SID[16] akan digunakan pada proses testing.

Tabel 1. Inisialisasi Parameter Awal

HR_{min} dan ϵ kombinasi berpasangan sesuai dengan kolom yaitu: [HR_{min} , ϵ] menjadi [0.4, 0.6],[0.5, 0.5] dan [0.6, 0.4].

5.2 Testing Metode dan Analisa

Hasil proses estimasi menunjukkan bahwa kombinasi nilai parameter optimal untuk metode yang diusulkan adalah [$S_T=0.3$, $HR_{min}=0.6$, $\epsilon=0.4$, $\theta=15$, $\alpha=0.8$, $T=0.3$] untuk ROUGE-1 dengan nilai rata-rata 0.32511 dan SID[16] dengan kombinasi

parameter [$S_T=0.4$, $HR_{min}=0.4$, $\epsilon=0.6$, $\theta=10$, $\alpha=0.5$, $T=0.4$] dengan nilai rata 0.31886.

Tabel 2. Hasil Testing

Metode	Nilai Rouge-1
SHCclustering+ClusterOrder+SID [16]	0.31886
SHCclustering + ClusterOrder + SID + FrameSRL + LabelMaching	0.32511

Dari proses testing didapatkan nilai rata-rata ROUGE-1 adalah 0.32511 sehingga disimpulkan bahwa metode yang diusulkan mempunyai korelasi yang lebih baik dari SID[16].

5. Kesimpulan

Ringkasan yang dihasilkan oleh metode yang diusulkan memberikan hasil yang lebih baik dalam pemilihan kalimat representatif dibuktikan berdasarkan nilai maksimum ROUGE-1. Korelasi dari hasil ringkasan sistem dan ringkasan manual yang terdapat dalam dataset.

Metode yang diusulkan menghasilkan rata-rata nilai 0.32511 lebih besar dari nilai metode SID[16] dengan rata-rata nilai ROUGE-1 0.31886. Hal tersebut mengindikasikan bahwa metode metode yang diusulkan mampu memilih kalimat ringkasan lebih baik berdasarkan korelasi hasil ringkasan system dengan ringkasan secara manual dari segi kesesuaian unigram dibandingkan dengan metode SID.

Daftar Pustaka

- [1] Barzilay, R. and McKeown, K. R. (2005) "Sentence Fusion for Multi document News Summarization." *Computational Linguistics*, 2009. ICSAP 2009. International Conference on IEEE.
- [2] Schlesinger, Judith D. O'Leary. Dianne P. and Conroy, John M. (2008) "Arabic/English Multi-document Summarization with CLASSY—The Past and the Future". IDA/Center for Computing Sciences. Ma, Xiao-Chen., Yu, Gui-Bin., and a, Liang. (2009), "Multi-document Summarization Using clustering Algorithm." *Computational Linguistics*. IEEE.
- [3] Carbonell, Jaime G. dan Goldstein, J, (1998), "The Use of MMR and Diversity-Based Reranking for Reordering Documents and Producing Summaries" Proceedings of the 21st meeting of International ACM SIGIR Conference, 335-336.
- [4] Gupta, V.K. dan Siddiqui, T. J. (2012), "Multi-Document Summarization Using Sentence Clustering". IEEE Proceedings of 4 th International Conference on Intelligent Human Computer Interaction, Kharagpur, India, December 27-29
- [5] Sarkar, K. (2009), "Sentence clustering Based Summarization of Multiple Texts Document." *International Journal of Computing Science and Communication Technologies*.
- [6] Hammouda, K. M. and Kamel, M. S. (2004) "Efficient Phrase-Based Document Indexing for Web Document clustering." *IEEE Transactions On Knowledge And Data Engineering*.
- [7] Ge, S. S., Zhang, Z. and He H. (2011), "Weighted Graph Model Based Sentence clustering and Ranking for Document Summarization."
- [8] He, T., Li, F., Shao, W. Chen, J. and Ma, L. (2008), "A New Feature-Fusion Sentence Selecting Strategy for Query-Focused Multi-Document Summarization." *International Conference on Advanced Language*

- Processing and Web Information Technology. IEEE.
- [10] Gildea, D. and Jurafskyy, D. (2001) "Automatic Labeling of Semantic Roles." International Computer Science Institute.
- [11] Baker, Collin F. and More, C. J. F. and Lowe, John B. (1998), "The Berkeley FrameNet Project." International Computer Science Institute Berkeley
- [12] Bjorkelund, A., Bohnet, B., Hafdell, L. and Nugues, P. (2009) "Multilingual Semantic Role Labeling." Department of Computer Science, Lund University,
- [13] Bjorkelund, A., Bohnet, B., Hafdell, L. and Nugues, P. (2010), "A High-Performance Syntactic and Semantic Dependency Parser." Department of Computer Science, Lund University
- [14] Palmer, M., Gildea, D., dan Kingsbury P. (2005) "The Proposition Bank: An Annotated Corpus of Semantic Roles." Association for Computational Linguistics.
- [15] Lin, C. Y. (2004), "ROUGE: a Package for Automatic Evaluation of Summaries", In Proceedings of Workshop on Text Summarization Brances Out, Eds: Moens, M. F. dan Szpakowicz, S., Association for Computational Linguistics, Barcelona, hal. 74-81.
- [16] Suputra, I.P.G.H, Arifin, A.Z, Yuniarti, A. (2013),"Pendekatan Positional Text Graph Untuk Pemilihan Kalimat Representatif cluster Pada Peringkasan Multi-Dokumen".Jurnal Ilmu Komputer Universitas Udayana.