

Luthfi Ali, 2013 *Implementasi TF-ISF (Term Frequency – Inverse Sentence Frequency) dan Title Overlap untuk Mencari Kalimat-Kalimat Penting pada Sistem Peringkat Dokumen Berbahasa Indonesia*. Skripsi ini dibawah bimbingan Badrus Zaman, S.Kom, M.Cs dan Indra Kharisma Raharjana, S.Kom, M.T. Program Studi S1-Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Airlangga.

---

## ABSTRAK

Saat ini, *Information Retrieval* digunakan dalam pengelolaan dokumen di internet dan pengambilan informasi terhadap dokumen-dokumen tersebut. Namun *output* dari *Information Retrieval* tidak sepenuhnya sesuai keinginan pengguna. Jika *output* dokumen sangat banyak, hal ini akan memberatkan pengguna untuk memeriksa dokumen satu per satu. Oleh karena itu diperlukan sistem peringkat teks otomatis (*automatic text summarization*) yang digunakan untuk meringkas isi dokumen, sehingga pengguna cukup membaca ringkasannya saja.

Untuk mencari kalimat-kalimat penting yang dijadikan sebagai ringkasan dokumen, sistem menghitung bobot tiap kalimat yang merepresentasikan nilai kepentingan kalimat tersebut. Metode yang digunakan untuk menghitung bobot kalimat dalam skripsi ini adalah TF-ISF (*Term Frequency-Inverse Sentence Frequency*) dan *Title Overlap*. Tahap dalam sistem ini untuk menghasilkan ringkasan adalah melakukan *preprocessing* terhadap dokumen, menghitung bobot tiap kalimat menggunakan *Title Overlap* dan TF-ISF, dan pemilihan kalimat-kalimat penting berdasarkan bobot tertinggi. Dokumen *sample* yang digunakan berupa 50 artikel berita yang didapatkan melalui situs berita *online*.

Proses evaluasi dilakukan dengan membandingkan ringkasan hasil sistem dengan ringkasan yang dibuat oleh tiga orang peringkat. Dari proses membandingkan tersebut didapat nilai *F-Measures* yang merepresentasikan tingkat keberhasilan sistem dalam menghasilkan ringkasan. Evaluasi dilakukan pada metode *Title Overlap*, TF-ISF, gabungan antara *Title Overlap* dan TF-ISF, dan fitur *Auto Summarize* pada Microsoft Word 2007 sebagai pembanding. Dari penelitian ini diperoleh hasil evaluasi sistem dengan nilai *F-Measures* sebesar 0.44 yang lebih tinggi dari fitur *Auto Summarize* pada Microsoft Word 2007 dengan nilai 0.4. Dari ketiga metode yang dievaluasi, metode *Title Overlap* memiliki *F-Measures* terbesar yaitu 0.496.

**Kata Kunci:** *Information Retrieval, Automatic Text Summarization, TF-ISF, Title Overlap*

Luthfi Ali, 2013 *Implementation of TF-ISF (Term Frequency – Inverse Sentence Frequency) and Title Overlap to Search for Important Sentences on Indonesian Document Summarization System*. This thesis is under guidance of Badrus Zaman, S.Kom, M.Cs and Indra Kharisma Raharjana, S.Kom, M.T. Bachelor Degree of Information System, Science and Technology Faculty, Airlangga University.

---

## ABSTRACT

Nowadays, Information Retrieval is used on managing documents in the internet as well as retrieving information from those documents. However, not every output of Information Retrieval is correct. If the output is so many, it will be a problem for the user to check them one by one to verify whether it is correct or not. Therefore, an automatic text summarization system is necessary for this, as it can summarize the content of a document and will help the user to verify the document just by reading its summarization.

To get important sentences as the summarization of the document, the system needs to get the weight of all sentences which represent its importance value among other sentences in the document. The methods that is used in this thesis to count the sentence's weight are TF-ISF (Term Frequency-Inverse Sentence Frequency) and Title Overlap. To generate a summary in this system some phases is needed, that is *preprocessing* on document, counting the weight of all sentences using *Title Overlap* and TF-ISF, and sentences selection as summary based on the highest weight. Document samples in this thesis are 50 news articles which are obtained from online news sites.

The evaluation process is done by comparing the system's output with summarized articles made by three subjects. The result of the comparison is *F-Measures* value which represents how accurate the system to summarize a document. Evaluation is done on *Title Overlap*, TF-ISF, combination of both methods, and *Auto Summarize* feature of Microsoft Word 2007 as a comparator. The *F-Measures* value of the system is 0.44 which is higher that Microsoft Word 2007 with *F-Measures* value of 0.4. From all methods that is evaluated, it is know that *Title Overlap* has the highest *F-Measures* value which is 0.496.

**Keywords:** *Information Retrieval, Automatic Text Summarization, TF-ISF, Title Overlap*