

INTISARI

ANALISIS PEMODELAN TOPIK PADA *LATENT DIRICHLET ALLOCATION (LDA)* UNTUK SISTEM *AUTO-TAGGING* PADA BERITA *ONLINE*

oleh

Annisa Titi Ramadhani
15/379617/PA/16675

Tagging atau penandaan menjadi salah satu metode yang populer dalam melakukan anotasi dan mengorganisir konten. Salah satu konten yang paling sering dimanfaatkan dalam *tagging* adalah teks artikel berita. Teks artikel berita menjadikan sarana untuk mendapatkan informasi terbaru namun terkadang penulis konten artikel berita melakukan penandaan secara manual. Hal tersebut dapat menyulitkan penulis jika dilakukan pada kumpulan artikel berita dalam jumlah besar. Oleh karena itu, diperlukan metode penandaan secara otomatis atau *auto-tagging* yang dapat memudahkan dalam melakukan penandaan pada teks artikel berita.

Penelitian ini bertujuan untuk mengimplementasikan pemodelan topik dengan metode *Latent Dirichlet Allocation (LDA)*, yaitu salah satu metode yang dibuat oleh Blei untuk mencari topik-topik apa saja yang terdapat dalam kumpulan dokumen dalam jumlah besar. Penelitian ini melakukan pemodelan topik dengan LDA sebagai penandaan otomatis pada 100 teks artikel berita berdasarkan hasil model LDA yang didapatkan. Hasil dari model LDA berupa sekumpulan topik yang setiap topik terdiri dari sekumpulan kata kunci. Metode yang digunakan dalam penelitian ini sebagai penandaan otomatis dengan mengimplementasikan model LDA dengan mencari topik yang dominan pada setiap artikel berita sehingga dari metode ini didapatkan beberapa kata kunci dari hasil pendistribusian topik.

Pengujian dilakukan dengan validasi model LDA dan mendapatkan jumlah topik yang optimal sebanyak 16 topik. Evaluasi hasil dilakukan dengan menggunakan perhitungan *confusion matrix* untuk mengukur nilai akurasi, presisi, recall, dan *f1-measure*. Hasil penelitian ini mendapatkan nilai akurasi 49%, nilai presisi 46%, nilai *recall* 94,70%, dan nilai *f1-score* 62% terhadap hasil dari implementasi LDA sebagai *auto-tagging*.

Kata Kunci: *auto-tagging*, pemodelan topik, LDA, *Latent Dirichlet Allocation*, *Gibbs Sampling*, *Confusion Matrix*, *tagging*

ABSTRACT

ANALYSIS OF LATENT DIRICHLET ALLOCATION (LDA) TOPIC MODELING FOR AUTO-TAGGING SYSTEM ON ONLINE NEWS

oleh

Annisa Titi Ramadhani
15/379617/PA/16675

Tagging becomes a popular technique to annotate and organizing content. An example of content that is often used in tagging is news article text. News article text is used as a tool for getting the latest information. Nevertheless, news article content writers often tag the content manually. This can also be troublesome for writers if it is done on a large collection of articles. Therefore, this research generates an automatic tagging method that can make an easier way to tag the news article.

This research proposed implementing a topic modelling using Latent Dirichlet Allocation (LDA), a method developed by Blei to determine what kind of topics are in a large collection of documents. This research also doing a topic modelling with LDA as an auto-tagging on 100 news articles based on a result that generated from LDA model. The results are some topics in which each topic has its own keywords. The method in this research is used to discover the dominant topics of each news article so that this method can get the keywords of topic distribution.

In this research, the test is conducted by validate the LDA model and be able to get 16 topics as the most optimal number of topic. The LDA-tagging result is evaluated with confusion matrix method by measure the values of accuracy, precision, recall, and f1-measure. In this research, the result shows values of 49% for accuracy, precision with 46%, recall with 94,70%, and f1-score with 62% over the LDA-tagging result.

Keywords: *auto-tagging, topic modelling, LDA, Latent Dirichlet Allocation, Gibbs Sampling, Confusion Matrix, tagging*