

ABSTRAK

PEMODELAN RANTAI TOPIK BERITA BAHASA INDONESIA DENGAN MENGGUNAKAN *LATENT DIRICHLET ALLOCATION*

Oleh

Muhammad Huda Widodo
15/383242/PA/16902

Membaca berita adalah salah satu alternatif yang bisa dilakukan untuk mencari informasi di masa sekarang. Berita sekarang dapat diakses dengan mudah melalui internet. Diperlukan analisis dan identifikasi topik dari suatu berita untuk mengetahui topik apa yang ada dalam kumpulan berita. Pembuatan rantai topik dibutuhkan untuk mengidentifikasi dan menampilkan topik yang ada dalam periode waktu tertentu. Selain itu rantai topik juga diperlukan untuk mengidentifikasi masalah atau isu jangka pendek dan jangka panjang dari suatu data berita.

Dalam penelitian ini, penulis meneliti tentang pemodelan topik berita Indonesia. Studi kasus dari penelitian ini adalah portal berita Kompas.com. Analisis dan pemodelan topik dalam penelitian ini menggunakan metode *Latent Dirichlet Allocation*. Pengujian pemodelan topik menggunakan uji perplexity menghasilkan 10 passes sebagai jumlah *passes* terbaik. Selain itu pengujian parameter jumlah topik menggunakan pengujian log perplexity didapatkan kesimpulan bahwa 9 topik merupakan jumlah topik terbaik.

Hasil dari topik yang dihasilkan dalam pemodelan topik selanjutnya dibuat rantai topik dengan menggunakan *Jensen-Shannon divergence* dan *Kullback Leibler divergence*. Pembuatan rantai topik pada penelitian ini menggunakan parameter sliding window dengan nilai 2 dan 3. Selain itu parameter similarity cut yang digunakan dalam membuat rantai topik adalah 2.5 untuk *Jensen Shannon Divergence* serta 1.6 untuk *Kullback Leibler Divergence* dapat berhasil menemukan isu jangka panjang dan jangka pendek dari sekumpulan data berita.

Kata Kunci : *Latent Dirichlet Allocation, pemodelan topik, topic chain, Jensen Shannon Divergence, Kullback Leibler divergence, sliding window.*

ABSTRACT

TOPIC CHAIN MODELLING INDONESIAN NEWS USING LATENT DIRICHLET ALLOCATION

By

Muhammad Huda Widodo
15/383242/PA/16902

Reading news is one alternative that can be done to find information in the present. News can now be accessed easily via the internet. Analysis and identification of topics from a news are needed to find out what topics are in the news collection. The topic chain making is needed to identify and display topics that exist within a certain time period. In addition, the topic chain is also needed to identify short-term and long-term issues or issues from a news data.

In this study, the author research about topic modeling of Indonesian news. The case study of this study is the Kompas.com news portal. Analysis and topic modeling used in this study are the Latent Dirichlet Allocation method. The topic modeling test using the perplexity test produced 10 passes as the best number of passes. In addition, testing the number of topic parameters using the perplexity log test, concluded that 9 topics were the best number of topics.

The results of the topics produced in the topic modeling are made into topic chains using Jensen-Shannon divergence and Kullback Leibler divergence. The topic chain making in this study uses sliding window parameters with values of 2 and 3. The similarity cut parameters used in making the topic chain are 2.5 for Jensen Shannon Divergence and 1.6 for Kullback Leibler Divergence can successfully find long-term and short-term issues from a set of news data.

Keywords : *Latent Dirichlet Allocation, pemodelan topik, topic chain, Jensen Shannon Divergence, Kullback Leibler divergence, sliding window.*