IMPLEMENTASI *FOCUSED CRAWLING* DENGAN *TOPIC-SPECIFIC*, *MULTINOMIAL NAIVE BAYES*, DAN *BREADTH FIRST SEARCH* GUNA PENGUMPULAN DATA *MEDIA MONITORING* GEOPARK CILETUH

TUGAS AKHIR

CLARA VELITA PRANOLO

1162001016

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK DAN ILMU KOMPUTER

UNIVERSITAS BAKRIE

JAKARTA

2021

IMPLEMENTASI *FOCUSED CRAWLING* DENGAN *TOPIC-SPECIFIC*, *MULTINOMIAL NAIVE BAYES*, DAN *BREADTH FIRST SEARCH* GUNA PENGUMPULAN DATA *MEDIA MONITORING* GEOPARK CILETUH

TUGAS AKHIR

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer

CLARA VELITA PRANOLO

1162001016

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK DAN ILMU KOMPUTER

UNIVERSITAS BAKRIE

JAKARTA

2021

Halaman Pernyataan Orisinalitas

Tugas Akhir ini adalah hasil karya saya sendiri, dan
semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama	:	Clara Velita Pranolo
NIM	:	1162001016
Tanda Tangan	:	
Tanggal	:	15 April 2021

Halaman Pengesahan

Tugas akhir ini diajukan oleh :

Nama	: Clara Velita Pranolo
NIM	: 1162001016
Program Studi	: Teknik Informatika
Fakultas	: Teknik dan Ilmu Komputer
Judul Tugas Akhir	: Implementasi Focused Crawling dengan Topic-Specific, Multinomial Naive Bayes, dan Breadth First Search guna Pengumpulan Data Media Monitoring Geopark Ciletuh

Telah berhasil dipertahankan dihadapan Dewan Penguji sebagai persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

DEWAN PENGUJI

Pembimbing : Guson P. Kuntarto, S.T., M.Sc. (..... 2021-04-25 GPK)

Penguji 1 : Yusuf Lestanto, S.T., M.Sc. (..... 15.04.2021 .....)

Penguji 2 : Irwan Prasetya Gunawan, ST, MEng, PhD. (.....)

Ditetapkan : Jakarta

Tanggal : 15 April 2021

Ungkapan Terima Kasih

Puji Syukur dipanjatkan kepada Allah SWT atas petunjuk, rahmat dan hidayah-Nya, sehingga terselesaikannya Tugas Akhir ini yang berjudul "Implementasi *Focused Crawling* dengan *Topic-Specific*, *Multinomial Naïve Bayes*, dan *Breadth First Search* guna Pengumpulan Data *Media Monitoring* Geopark Ciletuh" sebagai salah satu syarat dalam mencapai gelar Sarjana Komputer di Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Bakrie.

Penelitian dan penyusunan Tugas Akhir ini tidak akan terwujud tanpa dukungan serta bantuan dari berbagai pihak. Oleh sebab itu, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Kedua Orang Tua, Ibu Amalia Virnawati dan Bapak Tato Pranolo terima kasih untuk kasih sayang, kerja keras, pengorbanan, pengertian, semangat, serta dukungan yang tidak pernah ada hentinya diberikan untuk penulis dalam menggapai cita-cita sehingga terselesaikannya Tugas Akhir ini.

2. Bapak Guson P. Kuntarto, S.T., M.Sc. selaku dosen Pembimbing Tugas Akhir, yang selalu dengan sabar membimbing saya, selalu dapat meluangkan waktu, memberikan nasihat serta dukungan selama proses penelitian dan penyusunan Tugas Akhir ini.

3. Bapak Yusuf Lestanto, S.T., M.Sc. selaku dosen pembahas Tugas Akhir yang telah memberikan masukan dan perbaikan dalam penyusunan Tugas Akhir ini

4. Bapak Irwan Prasetya Gunawan, ST, MEng, PhD. selaku dosen pembahas Tugas Akhir serta dosen Pembimbing Akademik yang telah memberikan masukan, nasihat dan membimbing penulis dalam melangkah selama di perkuliahan sampai dengan di Tugas Akhir ini.

5. Bapak Berkah I. Santoso, S.T., M.T.I. yang pernah menjadi dosen Pembimbing Akademik yang telah memberikan masukan dan membimbing di semester awal perkuliahan.

6. Seluruh Bapak dan Ibu dosen Program Studi Informatika yang telah mendidik dan memberikan ilmu sehingga bertambahnya pengetahuan penulis yang tentunya sangat membantu dalam proses Tugas Akhir ini.

7. Mutiara Julia Ifra, Pilipus Delevia Vegas, Dezan Andhika, Hafiz Kurnia Aji, dan Fitrah Cahya yang selalu memberikan dukungan, pengertian dan baik selama diperkuliahan. Khususnya Mutiara Julia Ifra yang selalu sabar mendengarkan keluh kesah, memberikan dukungan supaya penulis dapat percaya diri menyelesaikan Tugas Akhir ini.

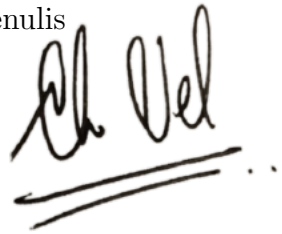
8. Teman-teman Informatika 2016, yang telah bersama selama 4 tahun.

9. Seluruh pihak Program Studi Informatika Universitas Bakrie yang telah memberikan ilmu dan pembelajaran serta pengalaman yang sangat bermanfaat bagi peneliti selama di perkuliahan.

Meskipun menyadari bahwa penelitian ini masih jauh dari kata sempurna, peneliti berharap bahwa penelitian dan Tugas Akhir ini dapat memberikan manfaat dan berguna bagi kalangan pendidikan, khususnya Informatika.

Jakarta, 15 April 2021

Penulis



Clara Velita Pranolo

Halaman Pernyataan Persetujuan Publikasi

Sebagai sivitas akademik Universitas Bakrie, saya yang bertanda tangan dibawah ini:

Nama	: Clara Velita Pranolo
NIM	: 1162001016
Program Studi	: Teknik Informatika
Fakultas	: Teknik dan Ilmu Komputer

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bakrie **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

Implementasi Focused Crawling dengan Topic-Specific, Multinomial Naive Bayes, dan Breadth First Search guna Pengumpulan Data Media Monitoring Geopark Ciletuh

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Bakrie berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta untuk kepentingan akademis.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di	: Jakarta
Pada tanggal	: 15 April 2021

Yang Menyatakan,

Clara Velita Pranolo

IMPLEMENTASI *FOCUSED CRAWLING* DENGAN *TOPIC-SPECIFIC*, *MULTINOMIAL NAIVE BAYES,* DAN *BREADTH FIRST SEARCH* GUNA PENGUMPULAN DATA *MEDIA MONITORING* GEOPARK CILETUH

Clara Velita Pranolo

ABSTRAK

Sebagai bagian dari *UNESCO Global Geopark* (UGG), wisata Geopark Ciletuh memiliki praktisi *Public Relations* (PR) yang mempunyai tujuan dalam hal membangun, mengembangkan, dan mempertahankan reputasi serta citra dari kawasan wisata Geopark Ciletuh. Untuk mencapai tujuan tersebut, praktisi PR melakukan *media monitoring* dengan mencari dan memilih berita dari berbagai sumber *media online*, lalu mengumpulkannya kedalam kliping berita untuk diidentifikasi dan dianalisis. Dalam *media monitoring*, proses ini disebut *data back-end*. Penelitian ini mengusulkan *focused crawling* untuk diimplementasikan pada *data back-end media monitoring* Geopark Ciletuh supaya proses pengumpulan data lebih cepat. *Focused crawling* diimplementasikan dengan menggunakan tiga metode yaitu metode *crawler* dengan Algoritma *Breadth First Search* (BFS) untuk mendapatkan URL berita yang lebih banyak, metode *distiller* dengan *Topic-Specific Weight Table* dan *Page Relevance* untuk fitur parameter dataset, serta metode klasifikasi dengan *Multinomial Naïve Bayes* untuk menentukan berita yang relevan. Hasil penelitian dengan algoritma BFS dapat melakukan *crawling* sebanyak 470 URL untuk Detik dan 290 URL untuk Kompas. Sedangkan dalam menentukan berita yang relevan akurasi yang didapatkan model *Multinomial Naïve Bayes* dengan *Page Relevance* yaitu 83.46% untuk *dataset* Detik, 89% untuk *dataset* Kompas dan diatas 88.16% untuk kedua gabungan *dataset* Detik dan Kompas.

Kata Kunci : *Web Crawling, Focused Crawling, Breadth First Search, Topic-Specific Weight Table, Page Relevance, Multinomial Naïve Bayes*

IMPLEMENTATION OF FOCUSED CRAWLING USING TOPIC-SPECIFIC, MULTINOMIAL NAIVE BAYES, AND BREADTH FIRST SEARCH FOR MEDIA MONITORING DATA COLLECTION GEOPARK CILETUH

Clara Velita Pranolo

ABSTRACT

As part of the UNESCO Global Geopark (UGG), Ciletuh Geopark has Public Relations (PR) practitioners who has the purpose to building, developing, and maintaining the reputation and image of the Ciletuh Geopark tourist area. To achieve this purpose, PR practitioners conduct media monitoring by searching and selecting news from various online media sources, then collecting them into news clippings to be identified and analyzed. In media monitoring, this process is called backend data. This study proposes focused crawling to be implemented in the back-end data of the Ciletuh Geopark media monitoring so the data collection process more effectively. Focused crawling is implemented using three methods, the crawler method with the Breadth First Search (BFS) algorithm to get more news URLs, the distiller method with the Topic-Specific Weight Table and Page Relevance for the feature parameter dataset, and the classification method with Multinomial Naïve Bayes for determine relevant news. The results of this study that BFS Algorithm crawl 470 URLs on Detik dataset and 290 URLs on Kompas dataset. The result also showed that Multinomial Naïve Bayes using the Page Relevance feature. The result also showed that Multinomial Naïve Bayes classify URLs based on Page Relevance with accuracy 83.46% on Detik dataset, 89% on Kompas Dataset and above 88.16% on combination of Detik and Kompas dataset.

Keywords : *Web Crawling, Focused Crawling, Breadth First Search, Topic-Specific Weight Table, Page Relevance, Multinomial Naïve Bayes*

Daftar Isi

Daftar Tabel

Daftar Gambar

Daftar Skrip

Daftar Singkatan

Singkatan	**Penjelasan**
PR	*Public Relations*
URL	*Uniform Resource Locator*
UGG	UNESCO *Global Geopark*
UNESCO	*United Nations of Educational, Scientific, and Cultural Organization*
TSWT	*Topic-Specific Weight Table / Topic-Specific*
BFS	*Breadth First Search*
WNRR	*Word that Not Represent Receipt*