

INTISARI

SISTEM PENCARI BERITA DENGAN MENGGUNAKAN *FOCUSED* *WEB CRAWLER* DAN ALGORITME *BREADTH-FIRST SEARCH*

Oleh

Ghiffari Agsarya Arlin

14/364209/PA/15944

Data berita sangat dibutuhkan dalam beberapa penelitian sebagai objek penelitiannya. Data berita dapat diperoleh dengan mudah dari situs berita *online* dengan cara mengunjungi situs berita yang diinginkan dan mengambil berita satu per satu. Dikarenakan data berita yang dibutuhkan jumlahnya banyak, maka cara tersebut tidak efektif, sehingga dibutuhkan suatu program untuk memudahkan proses pencarian data berita.

Sistem pencari berita yang dibuat pada penelitian ini menggunakan metode *keyword focused web crawler* dan algoritme *Breadth-First Search*. Metode ini melakukan proses pencarian berita sesuai dengan masukan *keyword* dengan memanfaatkan fitur *search* bawaan dari situs berita *online* yang dapat ditemukan pada situs berita yang dijadikan sebagai *seed Uniform Resource Locator* (URL). Metode ini akan membantu *crawler* dalam pencarian informasi yang paling relevan berdasarkan kata kunci pada situs berita *online* tanpa benar-benar melakukan *crawling* terhadap banyak *link* yang tidak relevan. Relevansi yang didapatkan sangat tergantung pada fitur pencarian tiap situs berita, apakah mengandung *keyword* yang dimasukkan atau tidak.

Hasil dari penelitian ini adalah menyatakan bahwa performa dari *keyword focused web crawler* lebih baik dari *web crawler* biasa dalam segi waktu *crawling* dan hasil pencarian. Dengan menggunakan parameter yang sama yaitu 10 *keyword*, 50 parent, dan tingkat kedalaman 3, antaranews.com mencatatkan waktu tercepat yaitu 49 detik, sementara detik.com dengan waktu 1 menit 35 detik, dan tirto.id dengan waktu 12 menit 10 detik. Rata-rata banyaknya berita yang berhasil di-*crawling* pada detik.com adalah 60 berita, antaranews.com dengan 72 berita, dan tirto.id dengan 171 berita. Rata-rata link yang diambil dari link ditawarkan pada antaranews.com adalah 14,43%, detik.com sebesar 14,43%, dan tirto.id sebesar 15,68%.

Kata Kunci: *Web Crawler, Keyword Focused Web Crawler, Algoritme Breadth-First Search.*

ABSTRACT

NEWS SEARCH SYSTEM USING FOCUSED WEB CRAWLER AND BREADTH-FIRST SEARCH ALGORITHM

By

Ghiffari Agsarya Arlin

14/364209/PA/15944

News data is highly needed in several studies as the object of research. News data can be obtained easily from online news sites by visiting the desired news sites and retrieving news one by one. Due to a large amount of news data needed, this method is not practical; therefore, a program to facilitate the process of finding news data is needed.

The news search system in this study uses the keyword-focused web crawler method and the Breadth-First Search algorithm. This method performs the process of news searching according to keyword input by utilizing the default search feature of online news sites, which can be found on news sites that are made as seed Uniform Resource Locator (URL). This method will help crawlers in searching for the most relevant information based on keywords on an online news site without actually crawling many irrelevant links. The relevance obtained very much depends on the search feature of each news site, whether it contains keywords that are entered or not.

The result of this study is the performance of the keyword focused web crawler is better than web crawler in terms of crawling time and search results. By using the same parameters which are 10 keywords, 50 parents, and a depth level of 3, antaranews.com records the fastest time of 49 seconds, while detik.com has a time of 1 minute 35 seconds, and tirto.id with a time of 12 minutes 10 seconds. The average number of news that successfully crawled on detik.com is 60 news, antaranews.com with 72 news, and tirto.id with 171 news. The average link taken from the link offered at antaranews.com is 14.43%, detik.com is 14.43%, and tirto.id is 15.68%.

Keywords: Web Crawler, Keyword Focused Web Crawler, Breadth-First Search Algorithm.