

## INTISARI

### **DISTRIBUTED WEB CRAWLER PADA SISTEM PENGADAAN SECARA ELEKTRONIK BERBASIS JARAK JARINGAN**

Oleh

Eko Sutrisno

13/356398/PPA/04387

Proses e-proc yang diimplementasikan di Republik Indonesia bertujuan untuk menciptakan iklim pengadaan barang dan jasa yang efektif, efisien, transparan, terbuka, terjadi persaingan secara sehat, adil serta akuntabel. Sistem Pengadaan Secara Elektronik (SPSE) memiliki server yang berbeda dan terpisah secara geografis dan belum memiliki sistem agregasi pengumpulan data yang dapat diakses oleh publik.

Sistem *distributed web crawler* (DWC) dapat melakukan otomasi pengumpulan data, agregasi data dan pencarian informasi yang spesifik dari server SPSE yang homogen dan tersebar tersebut. Sistem DWC memerlukan metode pembagian tugas yang terbaik agar kinerja tiap *Crawler Agent* dapat bekerja optimal. Penelitian ini bertujuan untuk menguji metode pembagian tugas *Polling System* (PS), *First Come First Serve* (FCFS) dan Jarak Jaringan Terdekat (JJT) pada 5 buah *Crawler Agent* dengan cara membandingkan akumulasi halaman yang terunduh, akumulasi jumlah data yang terunduh dan akumulasi jumlah *timeout* tiap satuan waktu.

Dari hasil pengujian diperoleh kesimpulan bahwa metode pembagian tugas JJT memberikan kinerja lebih besar dalam hal akumulasi jumlah halaman yang terunduh per satuan waktu daripada metode PS dan FCFS. Metode JJT dapat menunda *timeout* pada proses *crawling* yang sedang berlangsung daripada metode PS dan FCFS.

**Kata kunci** : Distributed Web Crawling, SPSE, Jarak Jaringan Terdekat

## **ABSTRACT**

### ***DISTRIBUTED WEB CRAWLER ON ELECTRONIC PROCUREMENT SYSTEM BASED ON NETWORK DISTANCE***

**By**

Eko Sutrisno

13/356398/PPA/04387

The e-proc process which implemented in Republik of Indonesia aims to create procurement of goods and services climate that are effective, efficient, transparent, open, healthy competition, fair and accountable. Electronic Procurement System (SPSE) has distinct and separate servers so it need a method to collect the data.

Distributed web crawler (DWC) system can automate data collection, data aggregation, and search for specific information from scattered SPSE servers. DWC system needs the best task scheduling method so that the performance of each peer can work optimally. This research aims to test the task scheduling method of Polling System (PS), First Come First Serve (FCFS) and Nearest Distance First “Jarak Jaringan Terdekat” (JJT) in 5 Crawler by accumulation of downloaded pages, accumulation of downloaded data and accumulated timeout per unit time.

From the test result we concluded that the JJT task scheduling method give greater performance in terms of accumulation of the number of pages downloaded per unit time than the PS and FCFS method. JJT method can delay timeout during crawling than PS and FCFS.

**Keywords :** Distributed Web Crawling, SPSE, Nearest Distance First