

Abstrak

Web Crawler atau biasa juga disebut *Web Spider* adalah bagian dari sebuah *Search System* yang berfungsi untuk menelusuri halaman web dan akan menyimpan semua informasi yang ada di dalam page tersebut. Untuk setiap web yang telah dikunjungi *Web Crawler* akan mengambil dan menyimpan semua link URL yang ada di dalam web tersebut. Kemudian URL tersebut akan disimpan dalam sebuah queue dan akan ditelusuri lagi sesuai dengan urutan yang telah ditentukan.

Sebuah *web page* dapat mengalami perubahan konten sesuai dengan keinginan pemilik web. Oleh karena itu dibutuhkan proses *revisit page* untuk memperbaharui database. Selain itu proses *revisit* dapat digunakan untuk proses perhitungan ulang PageRank jika terjadi perubahan hubungan antar dokumen web.

Pagerank adalah salah satu cara perangkingan web yang populer dan digunakan oleh *Search Engine* yang populer saat ini yaitu google. *PageRank* menggunakan prinsip *link* antar *page* dalam proses perhitungan tingkat kepentingan web. Sebuah halaman web akan memiliki *PageRank* yang tinggi jika ada banyak halaman lain yang meletakkan *link* menuju halaman tersebut.

Kata kunci: *Web Crawler, Pagerank, Revisit*