Hướng phát triển cho cralwer là dùng Nutch

Ban đầu định lấy link trực tiếp ở các trag web media rồi đưa link download

Lấy title để biết tên bài hát nhưng ko phải trang nào cũng để nội dung lên title , nhiều khi các nội dụng được đưa vào các thẻ khác , vì vậy phải đếm vị trí của thẻ để biết được nơi lấy nội dung

Mỗi trang 1 cấu tạo khác nhau

Có trang down đc có trang ko

Có trang source play media được code bằng ajax

Thêm thẻ link vào đầu để lấy tag <httrack>

Sau chuyển thẻ vì ko thuộc tag html

Trang phimtogo thi co 1 the comment ko giong nhu nhung trang khac

Co them the dau tien la the cua trang , the thu 2 moi la the comment can tim