

ABSTRAK

Data Scraping adalah proses untuk melakukan ekstraksi data dari suatu sumber (Dokumen atau website) untuk kemudian digunakan kembali sesuai dengan kebutuhan. Pada aplikasi ini dokumen yang akan di Scraping adalah PDF (Portable Document Format), PDF di pilih karena memiliki kelebihan yaitu sangat baik untuk proses publikasi dokumen karena format PDF tidak bisa diedit secara langsung selain itu file PDF juga berukuran relatif kecil di banding format lainnya. Untuk mengekstraksi data dari file pdf digunakan library PHP yaitu PDF Parser, PDF parser adalah library PHP yang menyediakan berbagai tools untuk melakukan ekstraksi dan pengolahan data dari file PDF.

Kata Kunci :Data Scraping, PDF, PDF Parser



ABSTRACT

Data Scraping is the process of extracting data from a source (document or website) for later reuse as needed. In this application, a document that will be scraping is a PDF (Portable Document Format), PDF was chosen because it has the advantage of being very good for the publication of the document because the PDF format can not be edited directly in addition to the PDF file size is also relatively small compared to other formats. To extract data from pdf file use PHP library that PDF Parser, PDF parser is a PHP library that provides a variety of tools to perform the extraction and processing data from PDF files.

Keywords: *Data Scraping, PDF, PDF Parser*

