

ABSTRAK

Judul : IMPLEMENTASI PENGANALISA SEMANTIK UNTUK EKSTRAKSI DATA KTP
Nama : Jhonny Cristian
Program Studi : Teknik Informatika

Setiap warga negara Indonesia yang telah berumur minimal 17 tahun memiliki sebuah kartu identitas berupa Kartu Tanda Penduduk (KTP). Bagian depan pada KTP ini berisikan identitas pribadi dari setiap individu, informasi ini berupa NIK, nama, tempat/tanggal lahir, jenis kelamin, golongan darah, alamat hingga data berupa pekerjaan. Informasi tersebut dapat diolah kembali menjadi sebuah data untuk dimanfaatkan lebih lanjut, contohnya pada saat pendaftaran aplikasi sehingga user tidak perlu melakukan penginputan nama, alamat dan tanggal lahir kembali. Untuk memproses informasi tersebut dibutuhkan sebuah sistem yang akan memproses informasi berupa gambar KTP menjadi sebuah teks yang dapat diolah kembali nantinya.

Pada penelitian ini sistem yang digunakan untuk mengekstrak gambar ke teks adalah Tesseract Optical Character Recognition (OCR). Tesseract OCR mengubah gambar menjadi teks acak dan tidak dapat membedakan masing – masing isi dari setiap teks yang telah diekstrak. Maka dari itu Analisis Leksikal dan Analisis Semantik digunakan untuk menutupi kekurangan tersebut, sehingga teks yang telah diekstrak sebelumnya akan dikelompokkan sesuai dengan jenis datanya.

Kata kunci : KTP, Tesseract, *Optical Character Recognition* (OCR), analisis leksikal, analisis semantik

ABSTRACT

Title : Implementation of Semantic Analyzer for ID Card Data Extract
Name : Jhonny Cristian
Study Program : Computer Science

Every Indonesian citizen who is at least 17 years old has an identity card in the form of an Identity Card (KTP). The front of this ID card contains the personal identity of each individual, this information in the form of NIK, name, place/date of birth, gender, blood type, address to data in the form of occupation. This information can be reprocessed into data for further use, for example at the time of application registration so that the user does not need to input the name, address and date of birth again. To process this information, a system is needed that will process information in the form of an image of an ID card into a text that can be reprocessed later.

In this study, the system used to extract images to text is Tesseract Optical Character Recognition (OCR). Tesseract OCR converts images into random text and cannot distinguish the individual contents of any extracted text. Therefore, Lexical Analysis and Semantic Analysis are used to cover these deficiencies, so that the previously extracted text will be grouped according to the type of data.

Keywords : KTP, Tesseract, Optical Character Recognition (OCR), lexical analysis, semantic analysis