ABSTRACT

The use of local languages is increasingly eroded due to the education system in many schools due to lack of maintenance adopting local cultural values and traditional and it lulled by the things that are too smelled internationally without going through the filter so that the younger generation are now more frequently using Indonesian. To preserve Sundanese, the necessary collection of digital word. The collection of Sundanese digitally've done at the end of a project undertaken by Novihana Nurani Setiawan titled WordNet Lexical Database Development Sundanese. At the end of the project Sundanese collection is still done by inserting the words one by one as a result require a long time for the collection of words and it took a lot of users, while users enter words not limited to users of Sundanese alone. therefore built Sundanese Word Extraction Application. Sundanese Word Extraction Application is a useful application for collecting words on a web page wikipedia Sundanese and Indonesian. The word that has been collected will be compared with the online dictionary to find the meaning of each word. In making application Extraction Sundanese word will use the library PHP Simple HTML DOM. Words collected from crawling classified into words and words Indonesian Sundanese.

Keyowrd: crawler, sinonim set, sundanese, PHP, HTML Simple DOM