

The aim of this work is to implement a system designed to effectively extract big amounts of data from web pages. The system is able to extract data both from simple static pages as well as complicated web applications. Data extracted from the web can have all sorts of types and can be interconnected by relations having various cardinalities. Working with system is based on XML documents, into which the given task is described in a declarative way. The description of a task is based on using system components, which connected together are able to do desired functionality on general web page.