This paper describes a system for automatic cleaning of HTML documents, which was used in the participation of the Charles University in CLEANEVAL 2007. CLEANEVAL is a shared task and competitive evaluation of automatic systems for cleaning arbitrary web pages with the goal of preparing web data for use as a corpus in the area of computational linguistics and natural language processing. We try to solve this task as a sequence-labeling problem and our experimental system is based on Conditional Random Fields exploiting a set of features extracted from textual content and HTML structure of analyzed web pages for each block of text.