The goal of this thesis is the comparison of methods for analysis of structured data sources, such as emails or HTML pages. The work focuses on practical assessment of common characteristics of these documents, which can be used for analysis, data extraction and cataloging for subsequent use. The work also includes a sample implementation of a program for cataloging data from emails and tracing changes in online sources.