Workflow-based Integration Framework

We investigate workflow-based approaches for online integration of web data. In the online setting, data integration tasks have to be executed at run-time, e.g., interactive web applications such as mashups. Such an ad-hoc data integration requires to combine content from multiple services or sources dynamically within a very short time. Existing mashup approaches only support rather simple data integration tasks.

Our framework is based on previous work on mapping-based data integration (\textit{iFuice}). It consists of components for search query generation and online matching as well as for additional data transformation. We support interactive and sequential refinement of query results to iteratively improve the result quality by executing more elaborate queries when necessary. A script-based definition of mashups (integration workflows) facilitates the development as well as the dynamic execution of mashups. The script programmer can use powerful generic operators to execute and manipulate mappings and their results. Based on the generic iFuice infrastructure we develop several domain-specific integration systems. An exemplary online system (online citation system, OCS) combines bibliographic data to dynamically calculate citation counts for journals, conferences, and authors (see related work on citation analysis).