

Collecting and Organizing Web Content

Mira Dontcheva
University of Washington

Steven Drucker
Microsoft Research

Geraldine Wade
Microsoft

David Salesin
Adobe Systems and
University of Washington

Michael Cohen
Microsoft Research

To collect and organize Web content today a user must make bookmarks, print whole webpages, or copy and paste pieces of webpages into a document. We present a framework for assisting the user in managing personal collections of Web content. The user interactively selects the webpage elements of interest, and the system builds an extraction pattern for those elements that is used to automatically collect content from analogous pages. Every new page the user visits is matched with existing patterns, and if matching page elements are found, the user can add them to the summary database. Layout templates filter the database to compose summary views. The user can view the database contents at any time with any layout template.

