Customizing Multimedia Information Access

DANIELA RUS

Department of Computer Science, Dartmouth College, Hanover, New Hampshire ⟨*rus@cs.dartmouth.edu*⟩

DEVIKA SUBRAMANIAN

Department of Computer Science, Rice University, Houston, Texas ⟨*devika@cs.rice.edu*⟩

We consider research questions in the design and analysis of *customizable agents* that capture, access, and retrieve heterogeneous, distributed, multimedia information. By agents we mean autonomous decision-making programs that are transportable across the Internet; by customizable we mean programs that evolve automatically with changes in the information landscape, as well as programs that can be easily modified and assembled by users, according to their tasks.

The transportable, customizable agent paradigm is well suited for distributed processing of heterogeneous data for the following reasons.

(1) Agents intelligently integrate heterogeneous information in response to user queries, and adapt to changing environments.
(2) Transportable agents possess routing autonomy and can travel from machine to machine to process information locally, avoiding costly and unnecessary data transfers over congested networks.
(3) Customizable agents are easily synthesized in a task-directed manner out of a library of prefabricated modules with performance guarantees.

For information agents to effectively filter multimedia distributed sources, we advocate a modular "open" architecture, configuring agents as *discrete control systems* with both feedback (reactive) and feedforward (planning) components. Modularization supports the orthogonal processing of different data dimensions (sound, video, pictures, text) and permits exploiting domain constraints to enhance performance. The computations needed to search and retrieve information are modularized using *structure* that occurs naturally at multiple levels of representation granularity in the data. By structure we mean any regularity that can function as a high-level, possibly task-specific index into information. To construct an agent for a task, the designer needs to identify structural cues to the desired information (e.g., information about economic indicators is likely to be found in tables and graphs in the "Money and Investing" subsection of *The Wall Street Journal*). In general, task-relevant structure is not readily apparent in raw data and computation is needed to reveal it. More specifically, our customizable agents are sensor-computational circuits [Balcazar et al. 1992] organized as adaptive, discrete control systems constructed from two basic types of modules: data segmenters (active structure "revealers" or data transformers) and structure detectors (or filters, e.g., for shapes in figures, tables in text, and melodies in audio).

The hypothesis that we test empirically is that standard conventions used by humans for presenting information define the basis of a finite library of parametric detectors and segmenters. We have developed modules for automatically discovering layout structures in scanned-in images of paper documents. These include detectors for automatically discovering titles, sections, definitions, figures, figure captions, and tables [Rus and Subramanian 1993; Rus and Summers 1995]. We are now developing similar filters for video and audio data.

The layout filters for paper documents are geometric in nature and work by identifying patterns in the distribution of white space on a page. For example, the table filter identifies candidate columns by looking for vertical rivers of white space on the page. A lexical analysis is performed across each candidate column to ensure that the information represented is homogeneous. We have analyzed the table-detection filter both geometrically and probabilistically to derive guarantees on its performance and a measure of its robustness with respect to imperfections in tables (such as misalignment and missing records.) We show, for instance, that for the algorithm to label with probability $1 - \epsilon_v$ a given column as a candidate column separator in a table, the number of white spaces in that column must exceed $0.2n + 0.4\sqrt{\epsilon_v n}$ for a block of text with n rows. The user-specified vertical error tolerance $0 < \epsilon_v < 1$ indicates how tightly the user wishes the columns to be aligned. This parametric table detector (characterized by three parameters: vertical and horizontal alignment tolerance and a lexical homogeneity tolerance) has been empirically tested on thousands of articles from several Usenet news groups [Rus and Subramanian 1994]. We have consistently obtained high recognition accuracies with a 3% false positive and a 2.7% false negative rate.

The modular agent assembly framework supports the construction of sophisticated integration strategies for retrieval from heterogeneous media [Rus and Allan 1995]. Indices into the information environment are associated with each parametric detector. We use definition-indices, figure-indices, caption-indices, table-indices, and the like for layout classification. A query for a specific definition in a technical paper is handled by searching through all indices. The retrieved structure links textual explanations, figures, tables, and graphs associated with the requested definition, integrating information that might be physically scattered throughout the document. Similar index structures can be established for other media using detectors of motion continuity for video streams or specific sound patterns for audio data.

Each index type has a well-defined similarity measure that assesses how related two indexed items are. For instance, for text we use the cosine distance in the vector space representation, and for computing shape similarity in images in video, we use the Hausdorff metric. Such measures allow the design of retrieval strategies that use information in one medium to retrieve related information in another medium. For instance, to integrate a figure denoting the Carnot cycle with explanatory text in another physically distant source, we use word indices on known sources (identified through the references) to find related text structures, and figure-indices to find similar visual shapes.

The prefabricated parametric modules can be implemented as static programs. For distributed environments, it is more advantageous to organize them as *transportable* software programs. Our transportable agents navigate the Internet using reactive planning methods and the virtual yellow pages for guidance. The agents make autonomous routing decisions by sensing traffic on the network and determining whether sites are up or down. They maintain partial maps (yellow-page caches) of the information environment, updating them as they go. The effort invested in searching for a specific piece of information is recycled by caching routes to locations where useful data were found. This cache guides the development of navigation strategies in the future. Because the information landscape and task constraints change, the yellow-page cache is not a static entity; simple adaptation schemes keep it up to date.

To assemble an agent for a specific multimedia search task, a user (a) identifies structural cues to the desired information; (b) constructs a retrieval strategy based on the identified cues/indexed; and (c) configures a controller that implements the strategy by composing corresponding detectors and segmenters from a library of modules. Rather than requiring the end user to program an agent from scratch for every information request, we provide a framework in which it is easy to assemble (akin to Lego blocks) *customized* multimedia information-processing engines that: (1) autonomously travel to where the data are located to filter the data locally; (2) adapt to the ambient environment in which they operate; and (3) are configured at the task level from modules that exploit the specialized structures in each information medium.

REFERENCES

BALCÁZAR, J. L., DÍAZ, J., AND GABARRÓ, J. 1992. *Structural Complexity I*, Springer-Verlag.

RUS, D. AND ALLAN, J. 1995. Structural queries in electronic corpora. In *Proceedings of DAGS '95: Electronic Publishing and the Information Superhighway* (May).

RUS, D. AND SUBRAMANIAN, D. 1994. Information retrieval, information structure, and information agents. *ACM Trans. Inf. Syst.* (Submitted).

RUS, D. AND SUBRAMANIAN, D. 1993. Multi-media RISSC informatics: Retrieving information with simple structural components. In *Proceedings of the ACM Conference on Information and Knowledge Management* (Nov.).

RUS, D. AND SUMMERS, K. 1995. Using white-space for automated document structuring. In *Advances in Digital Libraries, LNCS 916*, N. Adam, B. Bhargava, and Y. Yesha, Eds.