Proceedings of HLT-NAACL 2003
Demonstrations , pp. 5-6
Edmonton, May-June 2003

DOGHED: A Template-Based Generator for Multimodal Dialog Systems Targeting Heterogeneous Devices*

Songsak Channarukul and **Susan W. McRoy** and **Syed S. Ali**
Natural Language and Knowledge Representation Research Group
Department of Electrical Engineering and Computer Science
University of Wisconsin-Milwaukee
{songsak,mcroy,syali}@uwm.edu

1 Introduction

This paper describes DOGHED (Dialog Output Generator for HEterogeneous Devices), a multimodal generation component which is a part of a dialog system that supports adaptation of multimodal content based on user preferences and their current device. Existing dialog systems focus on generating output for a single device that might not be suitable when users access the system using different devices. Multimedia presentation systems can be built that support several device types. However, most content presentation and layout is done off-line and defined at the document level.

Dialog facilitates the process of tailoring the interaction to the dynamically changing needs of the user. With support for dialog, a computer can regulate the pace at which users receive content, help focus the user's attention, and interleave actions that will help the system monitor (and adjust) to the user's understanding or satisfaction. Minimally, dialog systems should adapt the interaction to the user's ability to understand. However, dialog systems should also be able to adapt to the user's computing environment, because people access computers not only through traditional workstations and terminals, but also through personal digital assistants and cellular telephones. Each of these devices has a distinct set of physical capabilities, as well as a distinct set of functions for which it is typically used.

2 DOGHED

DOGHED is a template-based multimodal output generator for dialog systems that need to support heterogeneous devices. It enables dialog systems to create multimodal presentations for different devices in real-time. DOGHED extends YAG (Yet Another Generator), a template-based text realization system (Channarukul et al., 2000; Channarukul et al., 2001; McRoy et al., 2003) by providing a pre-defined set of multimodal templates. It employs JYAG (Channarukul et al., 2002), the Java implementation of YAG, to realize those templates for further display by appropriate browsers. It provides output in the Synchronized Multimedia Integration Language (SMIL) which can be presented on any SMIL player such as RealOne Player and X-SMILES (Pihkala et al., 2001); or any capable web browser.

DOGHED is a generic and domain-independent component for generating multimodal presentations in that it accepts a feature structure as input. These feature structures can also embed other feature structures to represent a more complicated input specification. Moreover, an application can specify content, user preferences, and device constraints (*e.g.*, multimedia capabilities, screen size and resolution, and network bandwidth) using this uniform formalism. Natural language can also be inserted by embedding a feature structure that calls for realization using one or more YAG's English syntactic templates.

3 Multimodal Templates

There are two types of multimodal templates. The first type is for generating individual SMIL tags and structures. The other type is more abstract and captures the semantics of multimodal presentation. A semantic template can be authored so that an application only needs to specify its selected content, intended presentation style, and other constraints. Output will then be generated to best suit the given style and constraints. For example, an application might want to compare two diagrams or items that are related. On large display devices, both items can be displayed side-by-side. However, on smaller display devices, like PDAs, such presentation is not possible; it would be better to display one item at a time and switch between the two items being compared.

*We acknowledge the financial support of the National Science Foundation (under grants IRI-9701617 and DUE-9952703), Wright State University, and Intel Corporation.

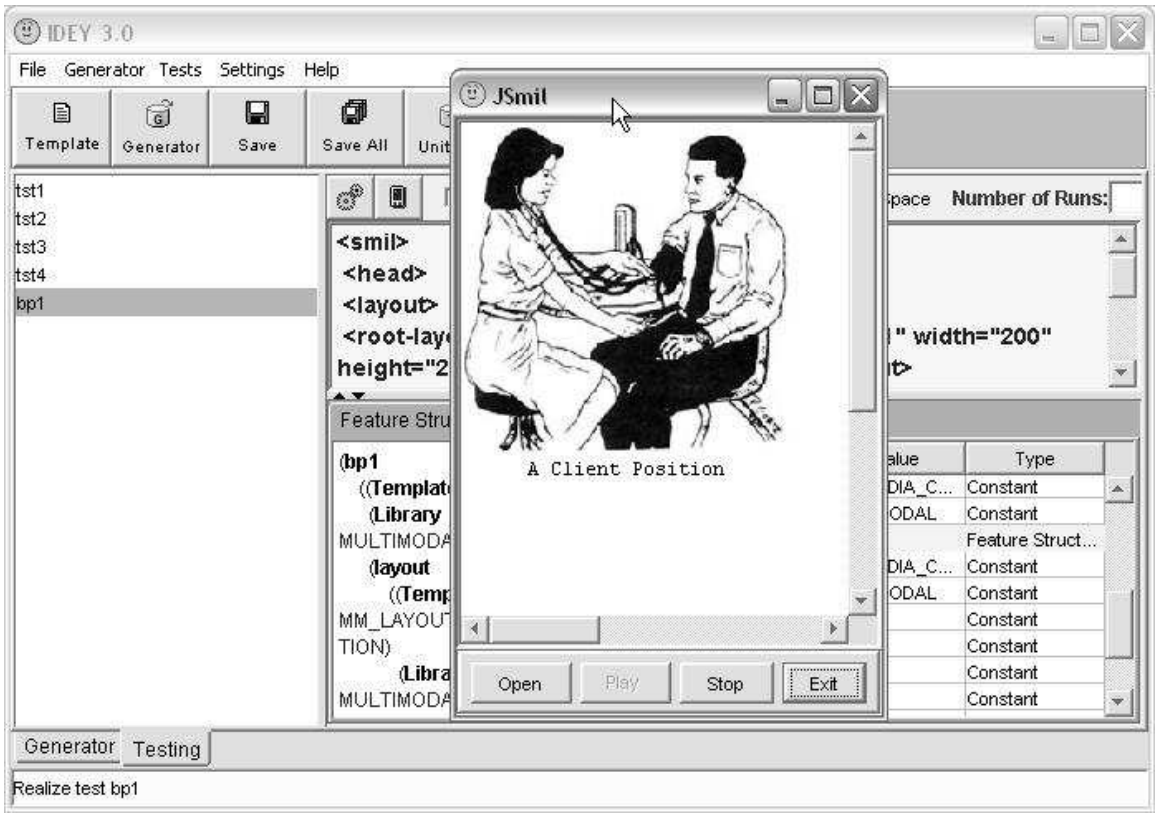

Figure 1: A Screenshot of IDEY with an Integrated SMIL Player.

4 Template Authoring Tool

In addition to the pre-defined multimodal template set, application developers can write their own templates to suit their needs. These new templates can be written from scratch or built on top of existing ones. We facilitate the task of template authoring by providing a graphical development environment. IDEY (Integrated Development Environment for YAG) provides support for authoring, testing, and managing templates (Channarukul et al., 2002). Its graphical interface reduces the amount of time needed for syntax familiarization through direct manipulation and template visualization. It also allows a developer to test newly constructed templates easily. Multimodal output can be immediately displayed and verified on an integrated SMIL player (Figure 1). Moreover, the interface also helps prevent errors by constraining the way in which templates may be constructed or modified. For example, values of slots in templates are constrained by context-sensitive pop-up menu choices.

References

Songsak Channarukul, Susan W. McRoy, and Syed S. Ali. 2000. Enriching Partially-Specified Representations for Text Realization using An Attribute Grammar. In *Proceedings of The First International Natural Language Generation Conference*, pages 163–170, Israel, June.

Songsak Channarukul, Susan McRoy, and Syed Ali. 2001. YAG: A Template-Based Text Realization System for Dialog. *Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems*, 9(6):649–659.

Songsak Channarukul, Susan W. McRoy, and Syed S. Ali. 2002. JYAG and IDEY: a Template-Based Natural Language Generator and Its Authoring Tool. In *Companion Volume to the Proceedings of the 40th Meeting of the Association for Computational Linguistics (ACL)*, pages 89–90, July.

Susan W. McRoy, Songsak Channarukul, and Syed S. Ali. 2003. An Augmented Template-Based Approach to Text Realization. *Natural Language Engineering*, 9(2):1–40.

Kari Pihkala, Niklas von Knorring, and Petri Vuorimaa. 2001. SMIL for X-SMILES. In *Proceedings of the Seventh International Conference on Distributed Multimedia Systems*, Tamkang University, Taipei, Taiwan, September.