Innovative Approach for Engineering NLG Systems: the Content Determination Case Study

Marco Fonseca, Leonardo Junior, Alexandre Melo, Hendrik Macedo

Departamento de Computação, Universidade Federal de Sergipe, 49100-000, São Cristóvão/SE, Brazil
marcos_ufs@yahoo.com.br, leonardobsjr@yahoo.com.br, asmelo10@hotmail.com, hendrik@ufs.br

Abstract. The purpose of Natural Language Generation (NLG) systems is that of automating the production of linguistically correct texts from a data source. Generators are usually built using ad-hoc software engineering practices, lacking a well-defined development process, standard software architecture, and the use of worldwide programming languages. This paper describes a new development approach that leverages the most recent programming languages and standards of modern software engineering to enhance the practical use of NLG applications. To show the practicability of the proposal, a content determination system, which accepts as input wrapped Web data regarding soccer championship results, was developed.

1 Introduction

Natural Language Generation (NLG) [13] is a conceptually consolidated technology. Past research has clarified many fundamentals issues and conceived solutions that are robust and scalable enough for practical use. Furthermore, opportunities for practical applications have multiplied with the information inundation from relevant Web content sources.

Unfortunately, NLG techniques remain virtually unknown and unused by mainstream and professional computing. This situation is probably due mainly to the fact that until recently, NLG was built using *ad-hoc* software engineering practices with no explicit development process and no standard software architecture. Reliance on special-purpose esoteric modeling and implementation languages and tools is another NLG issue. Every system is designed and implemented following specific domain complexities and needs and little has been done to change the portrayed situation. A good example is surface realization activity. Many realization components have been built based on different grammatical formalisms and theories used to describe NLG [8].

This work proposes an innovative approach to the development of NLG systems, in which the pipeline of text generation tasks work as a set of consecutive rule base for model transformation. Such methodology for building applications by applying transformations on models in different levels of abstraction was recently popularized