A UNL Deconverter for Chinese

Xiaodong Shi, Yidong Chen

Institute of Artificial Intelligence
School of Information Sciences and Technologies, Xiamen University
361005 Xiamen, China
{mandel, ydchen}@xmu.edu.cn

Abstract. This paper describes the internal working of a novel UNL converter for the Chinese language. Three steps are involved in generating Chinese from UNL: first, the UNL expression is converted to a graph; second, the graph is converted to a number of trees. Third, a top-down tree walking is performed to translate each subtree and the results are composed to form a complete sentence. Because each node is visited exactly once, the algorithm is of linear time complexity and thus much faster than the standard deconverter provided by the UNL center. A manual evaluation effort was carried out which confirmed that the quality of the Deconverter output was better than that of the standard deconverter.

1 Introduction

Although the UNL [1],[2] center provides a language independent generator [3] which can deconvert UNL expressions into any language provided that a UW dictionary, a set of deconversion rules, and optionally a co-occurrence dictionary are available for that language, that deconverter has a number of deficiencies: First, the deconversion rules are rather difficult to write because of the cryptic formats imposed by the deconversion specification. Second, although the power of the deconverter is claimed to be that of the Turing machine [4], its speed is rather slow and thus unsuitable for the main web application, embedded multilingual viewing of a UNL document that is one of the key goals of the UNL. Third, most importantly, the deconversion software is not open-sourced, so that fixing any bugs or introducing much-needed improvements is at the mercy of the UNL center, which has been rather lacking in technical support and in releasing new versions. So we think it is necessary to develop our own deconverter for Chinese. This paper describes such an endeavor. However, it should be noted that although we concentrate on generating Chinese from the UNL expressions, nothing in our deconverter is inherently related to Chinese, thus the deconverter is also language independent.

This paper is organized as follows: Section 2 will describe the main components of the deconverter and the algorithms involved. Section 3 will focus some issues in generation, especially those related to the Chinese language, and in Section 4 we will briefly discuss related work in the literature. In Section 5 we will give example uses of the deconverter and finally we will present the conclusions.

Universal Network Language: Advances in Theory and Applications.
Research on Computing Science 12, 2005, pp. 167–174.