学校编码：10384　　分类号_____密级_____
学　　号：20051302328　　UDC _____

厦 门 大 学

硕 士 学 位 论 文

基于浅层句法分析的翻译模板自动获取研究

Translation Templates' acquisition based on Partial parsing

林 哲 辉

指 导 教 师 ： 史 晓 东 教授
专 业 名 称 ： 计 算 机 应 用
论文提交日期 ： 2008 年　　月
论文答辩日期 ： 2008 年　　月
学位授予日期 ： 2008 年　　月

答辩委员会主席：_____

评　　阅　　人：_____

2008 年　月

厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年　　月　　日

厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1. 保密（　），在　　年解密后适用本授权书。

2.不保密（　）

（请在以上相应括号内打“√”）

作者签名：　　日期：　　年　　月　　日

导师签名：　　日期：　　年　　月　　日

摘　要

传统的 EBMT(Example-Based Machine Translation，基于实例的机器翻译)方法是建立在大规模的实例库基础之上的，存在着精确匹配率不高，模糊匹配时产生译文质量较差等缺点。利用翻译模板可以有效的解决翻译实例的数据稀疏问题、简化实例库的规模并提高实例匹配的精确率。因此，本文提出了基于两级翻译模板的方法，提出了两级翻译模板的体系结构并给出了定义，给出了从句子对齐的双语语料库自动抽取两级翻译模板的方法，并设计了利用模板库进行模板匹配的算法。

文中首先对机器翻译的重要技术尤其是基于 EBMT 的翻译方法进行了整理和分析，其次，概括了翻译模板提出的意义，提出了本文翻译模板的定义和研究策略。最后，在此基础上实现了两级翻译模板的抽取。

基于模板的翻译方法的核心问题是模板的抽取与匹配算法。本文在抽取翻译模板的过程中采用了浅层句法分析方法，主要运用了组块识别。浅层句法分析的优点是可以识别出确定性高的部分分析结果，减少句法分析中的歧义，从而降低句法分析的难度。具体方法是：在模板抽取过程中首先对翻译实例进行浅层句法分析，得到双语组块信息，并确定实例的框架，根据对齐约束提取短语翻译模板，在此基础上，提取出句子级的翻译模板。在模板匹配过程中首先对输入句进行组块分析，然后在系统模板库中检索匹配模板。检索过程中兼顾模板的结构信息及子块信息。

关键词：机器翻译　翻译模板　EBMT

Abstract

EBMT (Example-Based Machine Translation) systems are based on large scale example corpus in traditional, having the defect of low precision of matching. Translation template can solve the problem of data sparsity , large storage space and low matching precision of examples. Therefore, this thesis proposes one method which automatically extracts bi-level translation templates from sentence-aligned bilingual corpus, we propose its definition and matching algorithm. Finally, We proposed the matching algorithm as well.

Firstly, we analyze some important techniques of machine translation, especially the EBMT method. Secondly, we summarize the meaning of translation template, propose our definition and research strategy. Finally, we accomplish the extraction of bi-level translation templates.

The key point of EBMT method is translation templates' extraction and match algorithm. We used partial parsing in the process of extraction, chunking in general. The advantage of partial parsing is the robust result that can decrease the ambiguity of parsing. Concretely, partial parsing on the sentence-aligned bilingual corpus and get the chunking results, then we set the frame of the sentence, after extraction of phrase templates, we can get the sentence template. The matching module searches the most matching template for input sentence in the database, with the information of the chunking result and frame structure. The templates matching algorithm gets the searching result by using the key word in the sentence frame and its chunking template.

Key Words: Machine Translation; Translation Template; EBMT

目　录

Contents