Effective and Efficient Multilingual Information Access to Digital Collections

Jiangping Chen, Ph.D., Ryan Knudson, Zhongjie Chen, Min Namgoong
Department of Library and Information Sciences, University of North Texas

GOAL
The goal of this project is to determine the effectiveness and efficiency of applying Machine Translation (MT) technologies to enable Multilingual Information Access (MLIA) for digital collections.

OBJECTIVES
- Develop an MLIA service model based on MT for digital collections;
- Develop an open-source multi-engine MT (MEMT) system for metadata records translation;
- Develop a multilingual (English, Chinese, and Spanish) corpus of metadata records; and
- Evaluate the MLIA service model to determine which MT strategy achieves the best retrieval performance.

RESEARCH QUESTION
What is the effectiveness and efficiency of MLIA service applying machine translation to metadata records?

MODEL

Queries in Chinese or Spanish

DL or IR system

Queries in Chinese or Spanish

Chinese and Spanish Metadata Records

English Metadata Records

MT System

PHASES

Phase 1
- Establish Project Team and Train Research Assistants
- Acquire Metadata Records
- Pre-process Metadata Records
- Conduct MT with Selected Systems
- Generate Parallel Corpora
- Develop Test Queries

Phase 2
- Select and Install Retrieval System
- Configure Experimental System
- Select Metadata Records
- Configure Machine Translation Engine
- Perform MLIA Evaluation
- Conduct MLIA Experiments

Phase 3
- Upload Translated Metadata Records
- Develop and Test MEMT Strategies
- Document and Release system
- Analyze Evaluation Results
- Evaluate the Project and Disseminate Information

Phase 4
- Answer to the Research Question

Phase 5

Phase 6

EVALUATION
- Evaluating project success measured against project goal and objectives

GOALS
- Cost efficiency of the methodology
- Robust MLIA experimental environment
- MLIA experiments produce valuable findings
- MLIA experiments show high reliability and validity

OBJECTIVES
- Select open-source and low cost applications or services
- MLIA system is user-friendly
- MEMT system works well
- MLIA experiments show high reliability and validity
- Project website provides up-to-date information about project activities and is accessible by others
- Project findings and processes are shared with appropriate audiences
- Research question is answered
- Results are presented at relevant conferences and are shared in peer-reviewed publication

IMPACT
This project will have impact on three audiences:
- Non-English speakers
- Digital collection developers
- Researchers in Machine Translation and Cross-language Information Retrieval

CONTACT US
If you are interested in joining the team or learning more, please visit project website http://txcdk-iiia.unt.edu/MLIA/.