

An Investigation of Effective and Efficient Multilingual Information Access to Digital Collections

Jiangping Chen, Ryan Knudson, Min Namgoong

University of North Texas, Department of Library and Information Sciences

Purpose & Objectives

Explore the effectiveness and efficiency of applying machine translation (MT) to metadata records for multilingual information access (MLIA) services.

- Evaluate an MLIA model based on MT for digital collections;
- Develop multi-engine MT (MEMT) strategies for translating metadata records;
- Develop a multilingual corpus of metadata records for training the MEMT systems;
- Determine which MT strategy achieves the best MLIA performance.

MLIA Model

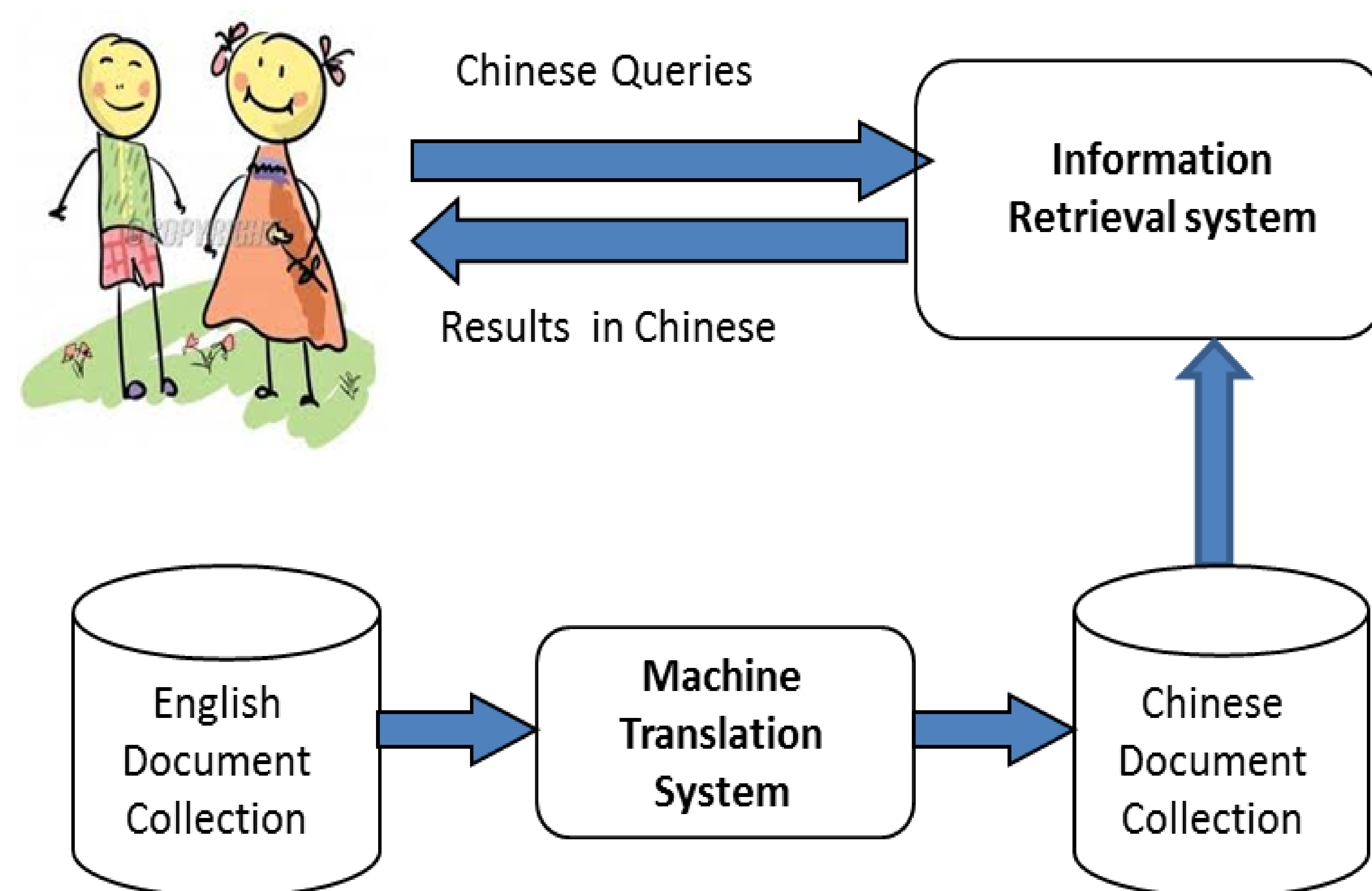


Figure 1. MLIA Based on Metadata Records Machine Translation

Project Procedures

- *Test Data Preparation*—Acquire 1 million metadata records from two digital collections; MT using two MT systems;
- *Multilingual Corpus Generation*—Develop a parallel corpus comprised of English, Simplified Chinese, and Spanish;
- *MEMT Experiments*—Use Moses to integrate MT results and linguistic resources to produce new translations;
- *CLIR Experiments*—Conduct Cross-Language Information Retrieval (CLIR) Experiments based on different MT results;
- *Evaluation* —Analyze results and measure effectiveness and efficiency of applying MT.

Planned Experiments

- English Baseline—monolingual retrieval
- Online System A—CLIR using 1st online MT system
- Online System B—CLIR using 2nd online MT system
- MEMT1—CLIR using MEMT that combines results of Systems A and B
- MEMT2—CLIR using MEMT1 + the multilingual corpora
- MEMT3—CLIR using MEMT2 + monolingual corpus

