Introduction

- Cross-language information retrieval (CLIR) has emerged as an important research area since the amount of multilingual web resources is increasing rapidly.
- The assumption behind CLIR translation mechanism is that queries are consistently used in one language, but such mechanism may not work well for bilingual users.
- In cultures where people use both Chinese and English, using mixed languages in spoken language and web searching is very common. A mixed language query is a search query including words mixed from two or more languages (e.g., Transformers 影评，Python 教程).
- Using Activity Theory as a methodological framework, this study examines how and why users employ mixed language queries in web searching.

Activities and Objectives

The activities of Chinese-English mixed language query users can be categorized into three types:

- Recreation searching: looking for a specific celebrity, place, book, movie, game, brand, product, and company originated in the Anglosphere.
- Work-related searching: varied by informants’ professions and their communities
  - searching for downloadable software or solutions to programming errors (software engineer)
  - searching for background information in Chinese about the western music materials (music librarian).
- Academic searching: looking for English translation of specific terminologies or how the Chinese research community translates or uses certain neologism.

The objectives of using Chinese-English mixed language queries can be summarized into three broad categories:

- To obtain high quality search results
- To retrieve webpages in Chinese to save reading time and effort
- To obtain webpages provided by certain communities (e.g., Chinese in Australia)

Research Questions

- In what activities do users employ Chinese-English mixed language queries? What are the objectives of these activities? What are the tools users employ in these activities?
- Why do users employ Chinese-English mixed language queries?

Research Method

- Qualitative semi-structured interviews were conducted online during October 2014 and December 2014 with 8 key informants who identified themselves as frequent users of Chinese-English mixed language queries and engaged in various professions (e.g., professor, software engineer, mechanical engineer, music librarian, and commercial consultant).
- All informants were native Chinese speakers living all over the world, including mainland China, Taiwan, Hong Kong, Japan, Australia, Europe, and the United States.
- Interview times ranged from 19 to 32 minutes with an average length of 26 minutes.

Tools

- Baidu
  - Used for recreation searching and looking for local resources in mainland China.
  - Some informants believed that Baidu did not index webpages in traditional Chinese characters and English published outside of mainland China, and thus failed to help them achieve completeness (or recall) in search results.
- Google
  - More useful for work-related searching and retrieving resources outside of mainland China, such as webpages in traditional Chinese characters.
  - Some informants considered it as less efficient in retrieving local Chinese resources.

Google's language setting function

- Not useful as the function was difficult to locate and the returned webpages containing English only.
- Some informants doubted if it only translated returned webpages to Chinese instead of retrieving pages containing Chinese.

Why Mixed Language Queries

- Why to use English terms
  - To limit search results to certain areas.
  - Unaware of corresponding English terms (e.g., PDF).
  - To retrieve resources provided by certain communities.

- Why to use Chinese terms
  - To direct search results to webpages in Chinese.
  - Did not know or were uncertain of the Chinese translations of certain English terms (e.g., error messages in programming).
  - No corresponding Chinese translations (e.g., MH370, Python, C++)
  - Various or no official Chinese translations (e.g., English personal names).
  - The English terms were used more widely in the Chinese communities (e.g., NBA, PSP, mp3).
  - The Chinese translations were hard to memorize or write.
  - The English terms were coined by Chinese.
  - Certain terminologies were too new to have Chinese translations (e.g., community informatics, social computing).

Future Research

- To collect more interview data and develop a thesaurus of English terms used in CLIR.
- To examine how other populations (e.g., Chinese learners) use English-Chinese mixed language queries in web searching.