# Chapter 99
# A Proposal to Study of Cross Language Information Retrieval (CLIR) System Users' Information Seeking Behavior

**YooJin Ha**
*Clarion University, USA*

## ABSTRACT

*There has been an enormous increase for information written in different languages by users from various backgrounds and disciplines. This chapter proposes a research design to examine multilingual information users' information behaviors when using a Cross Language Information Retrieval (CLIR) system. Development of a true CLIR is absolutely necessary so that the system would allow users to access information written in the user's languages of choice. Kuhlthau's Information Search Process (ISP) model was borrowed as a theoretical framework. Of particular concern are those users who want information represented by a language different than the users' original query or for those users who would like to retrieve additional information written in a second and/or third language or in a language which cannot be understood by them. This research is expected to yield a revised or new ISP model applicable to CLIR environments. It is expected that this study will also increase our understanding of CLIR users. The expected CLIR users include many of non-English speakers, especially users in developing countries who need this kind of CLIR system due to lack of materials in their own language. It is possible that the results of this research could inform CLIR system designers. The chapter is composed of purpose of study, literature review, theory, research questions, methodology, and discussion section. In the literature review section, pertinent research studies from information seeking behavior, cross language information retrieval, and general relevance studies are presented. Kuhlthau's ISP model is introduced in detail in the theory section. A possible application of Kuhlthau's ISP Model to the CLIR environment is presented in a table format. Research questions are developed from the literature reviews and Kuhlthau's model. Each research question, premises/assumptions, and its correspondent methodology are proposed in the methodology section. Limitations are discussed in the discussion section.*

## INTRODUCTION

Ranganathan (1957) states in his 2nd and 3rd laws of library science: "Every reader his book" (p. 80) and "Every book its reader" (p. 80). He would be a proponent of an open access system, which can help readers to "make discoveries" (p. 258) using all information, no matter what their formats or languages. Since the Internet began expanding globally in the 1990s, accessing various types and formats of information on the Web has become a daily practice for many information users in the world. People now depend more on the Web, digital libraries, and other information retrieval systems to search for information.

In the current Web environment, however, only a limited amount of resources can actually be usable to certain user groups. Although this may be related to access to particular subject areas in the invisible Web, the biggest reason for limitation to information is due to language differences. One of these examples includes information users in developing countries where there is a lack of information access to the materials in their own language (Ugah, 2007; Uhegbu, 2002; Etim, 2001) so people would need translation help from one to another language–mostly English to their own language (Parry, 2011).

According to Cybermetrics Lab, roughly 63% of the world's top 400 institutional repositories have non-English content (cited on Lederman, Warnick, Hitson, & Johnson, 2010, p. 126). Also as of 2011, the World Internet Usage data shows that 73.2% of the world's online populations are non-English speakers and it is roughly three-fourths of all Internet users. English speakers are only 26.8%, which is about one-third of all Internet users.

As Chowdhury (2003) affirmed, multilingual information retrieval has become a major challenge to gaining access to the prolific information on the Web (p.72). This challenge can be partitioned into several major components and this overview provides a structure for understanding the definition of and resolution for the symbolic representation of language differences between the query and the document and the environment in which the search occurs.

## Global Environment

Examples of the challenges confronting access to the written record will now be highlighted with special attention to the bibliographic surrogate which stands as the gateway to more extensive data such as reports, books, and journal information. Technological developments of the Web, the effects of globalization, and the international emphasis in academic scholarship are three contributors to creating demand for ethnically diverse peoples to study, work, and collaborate together across borders and continents.

Accordingly, today's information needs articulated by potential users of the Web and library databases expand the need for access to resources written in languages not known to the individual conducting the search. It is also possible, even likely in many situations the pertinent information being sought might be in a language where the query is in a different character set from the information being retrieved. It might be further assumed that English is used as a standardized language and that most people in search environments might possess some knowledge of written or spoken English. This is, however, a tenuous assumption. For example, suppose there is a branch of a Japanese company located in Korea, where the common language will be either Japanese or Korean (or possibly English). Here the search queries could be expressed in one of several languages spanning different characters and alphabets. Yet, the relevant information in the data base could be in those languages or other languages.

## Related Content

Document Clustering Using an Ontology-Based Vector Space Model
Ruben Costaand Celson Lima (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications  (pp. 1860-1883).*
www.irma-international.org/chapter/document-clustering-using-an-ontology-based-vector-space-model/198629

Modeling Domain Ontology for Occupational Therapy Resources Using Natural Language Programming (NLP) Technology to Model Domain Ontology of Occupational Therapy Resources
Ahlam F. Sawsaaand Joan Lu (2013). *International Journal of Information Retrieval Research (pp. 104-119).*
www.irma-international.org/article/modeling-domain-ontology-for-occupational-therapy-resources-using-natural-language-programming-nlp-technology-to-model-domain-ontology-of-occupational-therapy-resources/109664

Semantic/Fuzzy Information Retrieval System
Mounira Chkiwa, Anis Jedidiand Faiez Gargouri (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications  (pp. 1030-1050).*
www.irma-international.org/chapter/semanticfuzzy-information-retrieval-system/198587

Design of Context Cluster and Load Balanced Video Recommendation System in Cloud Computing
 Hanumantharaju R.,  Sowmya B. J.,  Aparna R.,  Shreenath K. N., K. G. Srinivasaand B. S. K. Sharanya (2022). *International Journal of Information Retrieval Research (pp. 1-18).*
www.irma-international.org/article/design-of-context-cluster-and-load-balanced-video-recommendation-system-in-cloud-computing/284489

The Effect of Stemming on Arabic Text Classification: An Empirical Study
Abdullah Wahbeh, Mohammed Al-Kabi, Qasem Al-Radaideh, Emad Al-Shawakfaand Izzat Alsmadi (2011). *International Journal of Information Retrieval Research (pp. 54-70).*
www.irma-international.org/article/effect-stemming-arabic-text-classification/64171