Human and Social Aspects of Information Seeking in Cross-Language Information Retrieval

Rowena Li
Bayside High School Library, USA

INTRODUCTION

With the rapid growth of computer and communication technology, a global interconnected information infrastructure is quickly constructing through the Internet. As the information is able to travel beyond geographical and spatial borders via the Internet, and more and more people around the world have gained access to global networks, the language boundaries have to be crossed in order to make the global communication possible. Statistics shows that 73.2% of Internet users are non-English speakers (MiniWats Marketing Group, 2012). However, English is still the leading language in global communication environment. English language dominates 55.5% of the websites (W3Techs.com, 2013). As the Internet has become one of the major communication mechanisms for information storage, retrieval, and dissemination, users need the ability to locate and retrieve information wherever, whenever and in whatever the language it has been stored. However, most of the search engines currently available can only provide monolingual information retrieval, which means that the retrieval can only be conducted in the same language as the query language. Cross-language information retrieval (CLIR) has become increasingly important to facilitate the effectiveness of information exchange among different languages. As a result, the study and development of tools and technology of cross-language information retrieval have gained greater attention over the past decade. While a lot of research has focused on the effectiveness of system functionality, few studies have examined information needs and social aspects related to cross-language information retrieval. This article aims to speculate the human and social aspects of cross-language information retrieval. It explores CLIR users' unique social and cultural contexts, their psychological and cognitive structures, and their distinctive relevance judgment. It examines in depth the barriers embedded in cultural, linguistic, and cognitive dimensions, which might hinder further advancement in cross-language information retrieval.

BACKGROUND

Information Seeking

Information has been traditionally viewed as a message transmitted from sender to receiver through a channel which may reduce or increase its ambiguity (Shannon & Weaver, 1949). In this view, information is external and objective, as well as structured and measurable. Dervin (1976, 1980, &1983) rejected this traditional approach and proposed her sense-making theory, which concerns the behavioral, cognitive and social aspects of information seeking. She introduced information seeking as a concept that an individual has to constantly "make sense" of his situation to move physically and cognitively across a gap in front of him through time and space. This model sees information seeking as systematic, subjective, situational, individual, cognitive, and holistic (Dervin & Nilan, 1986). In her own way, Dervin recognizes the similarity in information seeking process, as well as understands the uniqueness of individuals in information seeking situation. Saracevic, Kantor, Chamis, and Trivison (1988) associated information seeking process with users' *perception of the problem, intent for use of the information, internal knowledge state,* and *public knowledge expectations*. Wilson's model (1981) states that an individual's situation and social role, his/her psychological and cognitive states influence his/her context of information need. Other cognitive approaches to information seeking were also presented in the past two decades. Bates'

DOI: 10.4018/978-1-4666-5888-2.ch385

"berrypicking" model (1989) states that information seeking is not linear, but is constantly modified by the feedback occurred in information seeking process. Belkin's ASK (Anomalous State of Knowledge) model states that information seeking arises when an individual cannot identify the gap in his knowledge state (Belkin, Oddy, & Brooks, 1982). Taylor's (1991) situational theory believes that information seeking is not only "based on subject matter, but on other elements of the context within which a user lives and works" (p. 218). Kuhlthau's (1991) Information Search Process approach believes that individuals seek meaning rather than answers. Affective states are influenced by uniqueness of information situation and mood states. And personal involvement influences the information seeking process. Ingwersen's (1992, 1996) cognitive model states that cognition occurs in all stages of information seeking, and individuals experience a variety of cognitive modeling. Furthermore, Ellis's information seeking theory (1989), Schamber's relevance (1990), and Wilson's situational relevance (1973) have also influenced information seeking research.

In general, for the past two decades, the study of information has experienced a paradigm shift. It concentrates more on subjectivity (instead of objectivity) of information, on holistic (instead of atomistic) view of experience, and on internal cognition (instead of external behavior). And at the same time, users have become a part of information process and a significant factor in the study of information seeking and retrieval.

Information Retrieval and Cross-Language Information Retrieval

Information retrieval (IR) is "a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information" (Salton, 1968, p. v). It is the process of query formulation, matching, selection, evaluation and representation. The result is presented in the form of a ranked list in a descending order according to their relevance to the query. Information retrieval process is evaluated by the effectiveness and relevance of this returned list to the search query.

Most existing search engines provide only monolingual information retrieval; that is, the retrieved documents are in the same language as the query language. However, the information explosion on the Web leads to a strong demand for designing an effective cross-language information retrieval system, which permits users to find information regardless of language boundaries.

Cross-language information retrieval (CLIR) is a special case of information retrieval. It focuses on retrieving documents in languages other than query language. This process involves not only information retrieval, but also query translation to locate relevant documents. CLIR is defined as "Given a query in any medium and any language, select relevant items from a multilingual multimedia collection which can be in any medium and any language, and present them in the style or order most likely useful to the user, with identical or near-identical objects in different media or languages appropriately identified" (Fluhr et al., 1999).

CLIR Models and Approaches

The process of CLIR involves query formation, document pre-processing, index matching, document selection, document examination, document delivery, and relevance feedback (Oard & Diekema, 1998).

Of all the stages of information retrieval process, matching is the most important and problematic one. To better match a query to the indexed terms of preprocessed documents, four matching strategies have been identified. They are *cognate matching*, *query translation*, *document translation*, and *interlingual techniques* (Oard & Diekema, 1998).

Oard and Diekema (1998) indicated that *cognate matching* is the process of matching words from one language to another according to the words' spelling or pronunciation. This approach is used to find matches when the query words are not in the dictionary, but only under a condition - the two languages are in the same language group. *Query translation* is the most used strategy. The query is translated into target languages. Afterwards, the translation is used to match the documents for relevant retrieval. The advantage of query translation is its efficiency. However, since the query contains only a few words, the translation might lead to lexical ambiguity due to lack of context. *Document translation* is to automatically translate documents into query language. It involves online real-time machine translation system for users to navigate the Internet. However, it requires large system functionality. Since it is costly and time-consuming, it is only limited to smaller collections. *Interlingual techniques* convert