

1086-68-742

Christie Quaranta* (caquaranta@plymouth.edu), **Imanol Arrieta Ibarra**, **Eric Schwartz**
and **Elena Sizikova**. *Improving Cross-lingual Search Quality*.

Advancements in Information Retrieval (IR) focus primarily on increasing the speed and accuracy of search upon large data collections. One branch of IR is Cross-Language Information Retrieval (CLIR). With so many different languages across the world, it is imperative to be able to search in several languages effectively. Most CLIR methods involve machine translation; however automated translators are still very imprecise. Our research presents eight methods for cross-lingual search on the Shoah Foundation Institute Visual History Archive, as well as metrics to analyze the effectiveness of each method. We used the standard definitions of precision and recall, as well as a third metric, the F-measure, which is a weighted harmonic mean of the first two. First, a method in which the entire thesaurus is translated and searched is compared to another method SQTm to determine, which technique yields superior results. Moving forward with the more-effective SQTm, we improved recall and precision, and studied the methods for merits and downfalls. Recall-improving techniques presented involve expanding each search query to more “related” terms, while precision-improving methods utilize context-based filtering to narrow-down search queries by weeding out “irrelevant” ones. (Received September 11, 2012)