

Exercises for CLIR and CLQA

- What do we mean when we say an IR system is monolingual, multi-lingual or cross-lingual ?
- For the Machine Translation based approach for cross-lingual IR three different approaches are distinguished. Name them and briefly describe them.
- For the index translation-based approach of CLIR, why is ambiguity of index terms a problem ? Give some examples.
- Name a major difference between cross-lingual IR and cross-lingual QA.

Continued

- For the following questions translate each to English using Google translate, Bing Translator and one online translation system of your choice and merge the resulting three different translations into a single common one using a bag-of-words approach (following the idea of step 1 in slide 49):

„Welches Land wurde 1978 Fußballweltmeister?“

„Коя страна стана световен шампион по футбол за 1978 г.?“

„Mistä maasta tuli jalkapallon maailmanmestari vuonna 1978?“

„¿Qué país se proclamó campeón en los Mundiales de Fútbol de 1978?“

- Send the translated questions to a search engine. Do this for the individual translations and for the merged translation. Assess the top 5 snippets as „relevant“ or „non-relevant“. A snippet is „relevant“ if it contains an answer string.
- For those snippets which you assess as „non-relevant“ check the corresponding document whether it contains an answer string. Mark the document accordingly.