В наше время огромное количество информации доступно в электронном виде. Информационные системы, оперирующие большими объемами данных произвольной предметной области и успешно решающие различные прикладные задачи, становятся все более востребованными, как предприятиями и организациями, так и отдельными пользователями.

Информационный поиск(IR) представляет собой процесс извлечения релевантной информации среди огромного количества документов. Традиционные IR системы реализуются в основном для документов, написанных на одном языке, хотя интернет сам по себе является многоязычной информационной средой. По этой причине возникает языковой барьер между пользователем и доступной информацией, а также появляется необходимость в исследовании и разработке методов для повышения эффективности IR.

В большинстве случаев при поиске информации в интернете мы хотим, чтобы она была написана на нашем родном языке, однако такая информация не всегда является доступной. С учётом того, что большинство пользователей владеет 1 или несколькими иностранными языками, они могут быть также заинтересованы в поиске информации, написанной на других языках. Так появляется необходимость в многоязычном поиске(CLIR), целью которого является сопоставления запроса, написанного на одном языке, с документами, написанными на других языках. CLIR снимает языковой барьер, благодаря чему пользователи могут отправлять запросы, написанные на их родном языке, а получать документы на других языках и наоборот. Например, запрос на русском языке вернёт релевантную информацию на английском языке. Из-за быстрого развития интернет-технологий потребность в CLIR значительно растёт, поскольку данный тип поиска позволяет реализовать обмен информацией между различными языками, устранить лингвистическое несоответствие между предоставляемыми запросами и документами, которые извлекаются из информационной сети. В связи с этим CLIR приобрел большое значение, как в качестве исследовательской дисциплины, так и в качестве технологии, которая будет востребована на рынке.

В дополнение к проблемам, встречаемым при одноязычном IR, в CLIR добавляется ещё одна – проблема перевода. Однако в данном случае перевод будет отличаться от полнотекстового машинного перевода. Причиной этому является отсутствие необходимости быть удобочитаемым для человека, перевод должен просто максимально подходить для поиска соответствующих документов. В основе CLIR могут лежать следующие варианты реализации перевода: перевод запроса, перевод документов, перевод запроса и документов одновременно. Уже было опубликовано большое количество исследований по теме реализации CLIR. Многие вопросы, связанные с данной темой, также рассматриваются на различных конференциях, например, TREC, NTCIR, CLEF. Каждая из данных конференций охватывает определённые языки: TREC включает в рассмотрение испанский, китайский, немецкий, французский, арабский и итальянский языки; NTCIR включает японский, китайский и корейский языки, а CLEF - французский, немецкий, итальянский, испанский, голландский, финский, шведский и русский.

В дипломной работе сначала приводится описание подходов реализации поиска релевантных документов в одноязычной информационной среде. Затем выполняется анализ техник перевода, а также методов разрешения лексической многозначности для осуществления поиска в многоязычной информационной среде. Итогом проведенного в дипломной работе исследования является разработанное мобильное приложение, обладающее cross-language функциональностью при поиске релевантных документов в сети интернет.