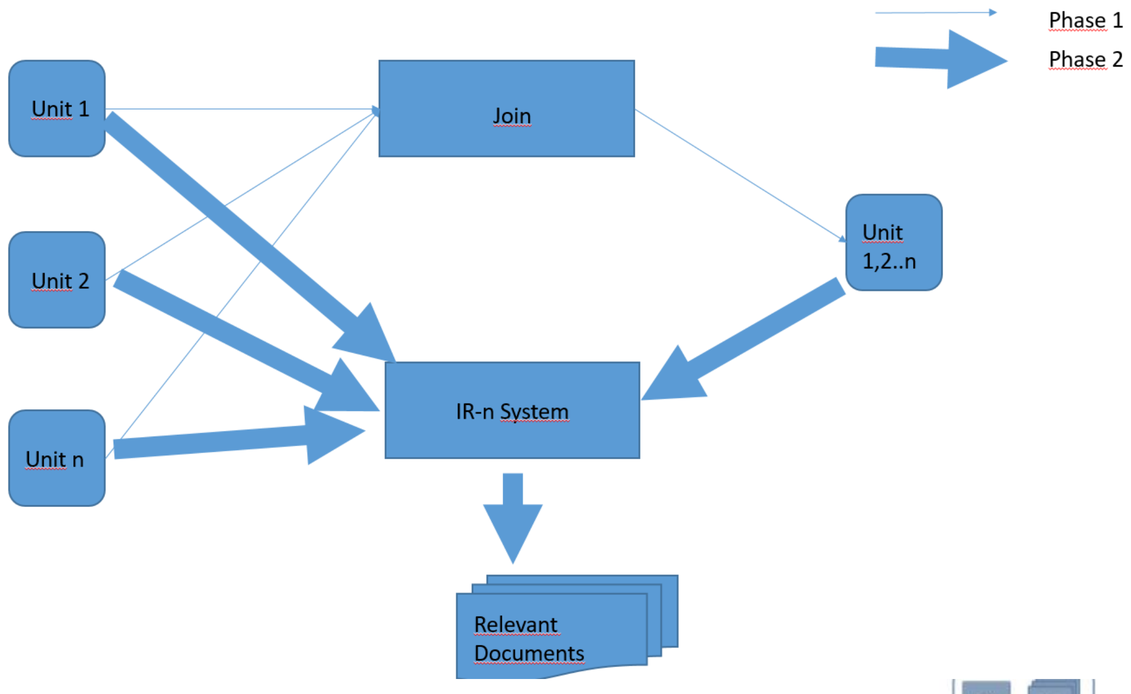
University of Alicante at the NTCIR-12: Mobile Click

Fernando Llopis Elena Lloret José M. Gómez

University of Alicante
Spain
{Fernando.Llopis@ua.es, elloret@dlsi.ua.es,, jmgomez@ua.es}

Abstract

This posterdescribes the first participation of processing natural language group of the University of Alicante in Mobile Click Task of NTCIR 12. Our approach is based on the combination of tools developed in our research group: IR-n, a passage retrieval system; COMPENDIUM, a summarization generator; and a new approach based on Principal Component Analysis, another type of summarizer. In our first participation we focused on the iUnit Ranking Subtask, although we have made an attempt on the iUnit Summarization Subtask.

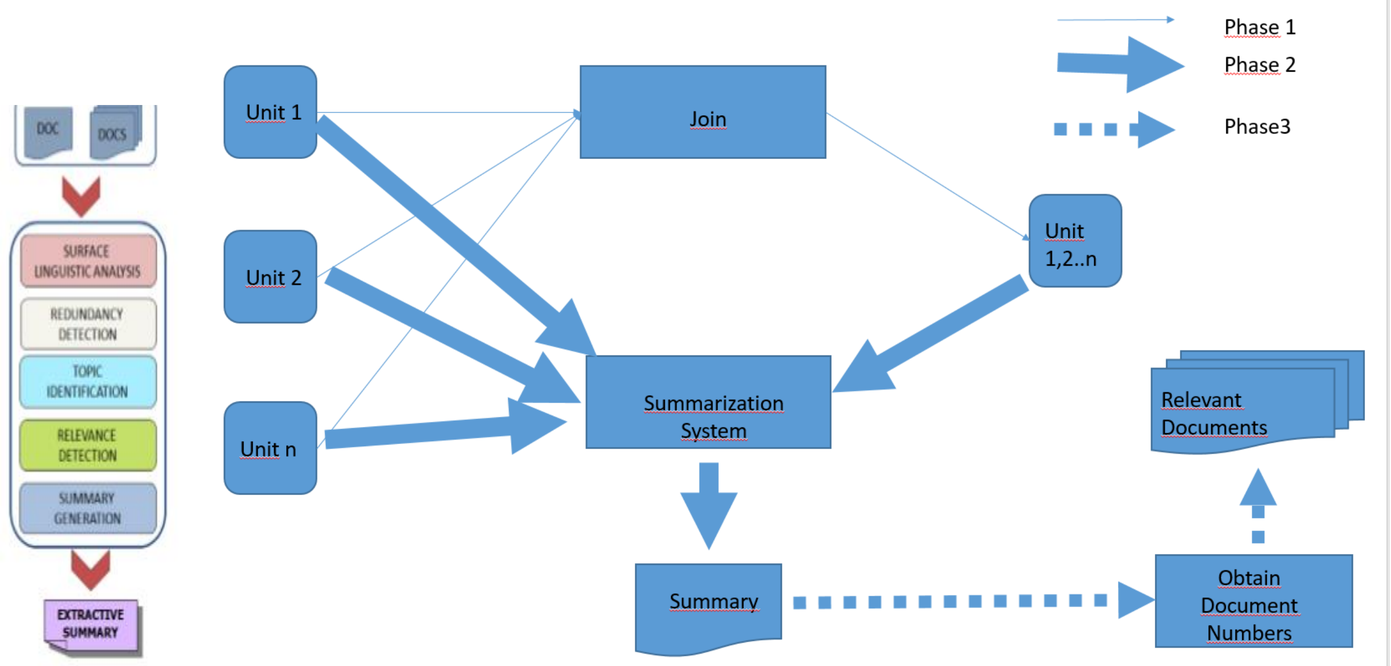


IR-n System Passage retrieval System

Passage Retrieval is an alternative to traditional document oriented Information Retrieval. These systems use contiguous text fragments (or passages), instead of full documents, as basic unit of information. IR-n system is a passage retrieval system that use groups of contiguous sentences as unit of information. Pass

Compendium Summarizer

COMPENDIUM is an automatic text summarisation tool that produces generic informative extracts from single or multiple documents. For the identification, selection and extraction of the most relevant information, different techniques are employed trough a pipeline of five stages.



		Q-measure		
		10%	20%	40%
Baseline (IR-n)	**0.8621**			
COMPENDIUM		0.8196	0.8291	**0.8403**
Semantic PCA-based		0.8101	0.8040	0.8064
COMPENDIUM+TE		0.82010	0.8218	0.8246

Table 1: Results with summarization systems alone

		Q-measure
Baseline (IR-n)	**0.8621**	
COMPENDIUM		**0.8403%**
COMPENDIUM + IRn		**0.8648%**

Table 2: Results for the combination IR-n+COMPENDIUM

	Q-measure
IRn + COMPENDIUM (*)	**0.9027**
TTTEC	0.9003
UHYG	0.8994
ORG	0.8975
RISAR	0.8972
RISAR	0.8962
IRn	0.8959
COMPENDIUM	0.8934

Table 3: Results for Test iUnit Ranking Subtask

RESULTS

iUnitRankingSubtask The results can be seen in Table 3. We obtained the best result with IRn+COMPENDIUM. However, this was not an official score because we could not finish the training on time. iUnitSummarizationSubtask We did not obtain a satisfactory results. This was due to the fact we did not have enough time to work on this task.

Future Work

As future work, we intend to work with the queries to obtain more information, so that we can improve our system's precision. Moreover, we need to investigate the effects of query expansion techniques over the intents. Finally, we are also trying to improve the ordering of sentences extracted by the summarization systems to improve the user experience at reading.