

**NUS Graduate School for Integrative Sciences and Engineering
Research Project Write-up**

Title of Project : Precise Retrieval and Question-Answering

Name of Supervisor : Chua, Tat-Seng

Contact Details: chuats@comp.nus.edu.sg

Short Description

Question answering (QA) aims to find exact answers to users' natural language queries, instead of ranked lists of documents as is done in current search engines. It is a major step towards information retrieval instead of document retrieval. Our QA system employs a pipeline structure that consists of several modules to get short and precise answers to users' questions. It searches for answers at increasingly finer-grained units of: (1) locating the relevant documents, (2) retrieving passages that may contain the answer, and (3) pinpointing the exact answer from candidate passages. The research focus of our work is three-fold. First, we search the Web for relevant context information to supplement the often inexact query. In particular, we perform semantic clustering of information retrieved from the Web to identify different sub-events and induce different facets of queries in supporting event-based QA. Second, in addition to density-based word matching, we employ discourse, semantic and dependency relations to perform passage retrieval at sentence level. This gives rise to a multi-resolution framework for relation-based precise information retrieval. Third, we develop a document concept lattice model together with definitional patterns and a human interest model to perform task-oriented summarisation. Our studies on the large-scale TREC-QA corpus demonstrate that our approaches are effective in performing factoid, list and definitional QA. Our system has been ranked consistently at second position over three years (2003-2005) in the public TREC-QA evaluations organised by NIST, USA. Our summarisation system also achieved top position in the DUC forum in 2005. Our technology has been licensed by industry to perform precise legal search.

Sub-projects of interests are: (a) multi-resolution relation-based framework for information extraction and QA; (b) QA on the Web; and (c) interactive QA.