Abstract

Good requirements have to be gathered for software development. Most of the requirements from the client of a proposed software system are available in informal or unstructured documents. Most often requirements are ill defined. Deficiencies in software requirements are mostly identified only after deployment. Most software requirements data available to software engineers are expressed in natural language & 90% of data are unstructured. Customers might not be able to provide all the requirements since they are not sure of what they want. Also they are unable to state the requirements due to incomplete knowledge of the applications functionality. This paper proposes methods to derive formal specifications from
informal or unstructured documents using techniques from natural language processing & text mining. The purpose of using natural language processing techniques is that sentences should be understood without human intervention and to understand the document in a way how human beings interpret & understand. The objective is to elicit the requirements from informal documents without using any strict template & to keep human intervention to the minimum. A parser or POS tagger is to be used to find the nouns, verbs. Transform free form text into an Intermediate Form (IF), Extract Entities & Relations then cluster Entities & Relations. Natural language transcripts contain lots of user requirements specified in their own language without any technicalities. For program development these informal requirements have to be derived, studied for their feasibility and then built into the software which would satisfy the requirements of the customer. Capture the requirements for the system and generate specifications from the captured requirements so that the software engineers and customers can understand them.

**References**

- Yu-shan Chang et al, Stanford University, Applying Name Entity Recognition to Informal Text.
- Marinos G Georgiades et al, IADIS International Conference Applied Computing, 2010, A Novel Methodology to formalize the requirements engineering process with the use of natural language.
- Carlos Mario Zapata, NAACL HLT 2010 Young Investigators workshop on Computational Approaches to Languages of the Americas, California, June 2010, Computational Linguistics for helping Requirements Elicitation: a dream about automated software development.

**Index Terms**

Computer Science

Information Technology
Keywords
Unstructured Documents  Information Extraction  Natural Language Processing  Text Mining

Requirements Engineering