

Income Tax Department

Request for Expression of Interest for development of Software tool for text mining

1. Objective of this EOI

The following document provides background information on the Income Tax department's (ITD) requirement for a software tool for text mining in various formats. This is only an expression of Interest to elicit the technical capabilities of the Respondents and to understand mature and established technologies and tools available. Based on the evaluation of responses, ITD will short list the respondents for issue of detailed RFP. ITD reserves the right to reject any proposals submitted in response to this EOI.

2. Overview of the requirement

One important function of ITD is survey and search operations wherein documents in electronic or physical form are impounded or seized by Income Tax authorities for verification. ITD undertakes this exercise as part of its enforcement function which is done by the Officers in Investigation Wing as well as Assessing Officers. At present verification of these documents is carried out manually. Method of record keeping and book keeping has undergone a paradigm shift in the last decade. Most businesses (corporate and non corporate) have switched over to computer based documentation for efficiency, speed and control over the affairs of their organizations. Often these are integrated with financial packages. Several documentation and accounting software are available in the market.

Most developed tax administrations use software tools for analysis and extraction of information from these electronic documents. ITD has also prepared a road map to computerize its functions in line with international best practices. One of the components of this is to develop a software based text mining tool to analyze the computerized documents that will assist in investigation process. ITD is looking for a text mining tool that will be able to mine the unstructured text and data in such electronic documents in any format (MS Word, PDF, spread sheets like lotus, excel, e-mails etc) and thereafter help the departmental officers to extract and analyse information from these documents for quick and effective investigation.

The proposed tool should be able to extract summaries of the words contained in the documents based on the parameters given by ITD.. The tool should be able to handle text and data in various document formats on 'as is where is' basis. The tool should possess all the functionalities of a standard text mining tool and should be able to analyze words, clusters of words used in documents, analyze documents and determine similarities between them or how they are related to other variables of interest for extraction of information. The tool should have the capability to be scaled up to analyse

multi-user, multi-location documents. The tool should be able to handle large documents which are structured or unstructured.

Interested Parties may respond to this proposal giving the full details of the tools developed by them explaining all the functionalities. ITD will test the tools offered by the Respondents which will include a proof of concept on some actual data.

3. For the attention of Vendors

ITD is inviting proposals from Respondents possessing software tools that can meet above requirements. This EOI is only a request to the Respondents to showcase the technical details and functionalities of the tools available with them. The Respondents should not include any financials whatsoever with their responses. It may be kept in mind that this is only an Expression of Interest and does not give any right and claim whatsoever to any Respondent nor creates any obligation on ITD for acceptance whatsoever.

Based on the evaluation, the Respondents will be short listed and a detailed RFP (Request For Proposal) will be sent to them. The RFP will have a detailed methodology of evaluation covering the technical as well as commercial aspects.

The EOI should cover the following

- Introduction of the Respondent
- Detailed overview and architecture of the text mining tool.
- Detailed note on the functionalities of the tools with the types of text and document formats that the software can handle.
- Capability regarding the scalability of the tool.
- Details of Previous installations of the tool including the name of the organization and the contact number of the references in the organization.

4. Submission of EOI

The EOI should be mailed to the following e-mail ashishabrol@incometaxindia.gov.in or delivered by post to the following address :

ASHISH ABROL
Jt. DIT (Inv) Spl.Cell
Room # 348, E2, ARA Center
Jhandewalan Extn.
New Delhi (India) 110055

5. Evaluation of the responses

A committee of the officers of the ITD will evaluate the responses received. The evaluation will be based on documents submitted and also a demonstration of the complete solution on a sample data in ITD premises either at Hyderabad or New Delhi. Only the Respondents willing to do a Proof of Concept at the ITD premises on the data supplied by the department may respond.