

SIB (103)/CCTNS/ML-CAS/87/NCRB-2011
Government of India
Ministry of Home Affairs
National Crime Records Bureau

East Block-7, R.K. Puram,
New Delhi-110 066
Date: 23rd October 2012

To

Nodal Officers (CCTNS Project)
All States / UTs

Subject: Development of police domain specific business dictionary in local language - reg

Sir,

This is with reference to the multilingual functionality in core application software (CAS). The software development agency (SDA) is in process of evaluating CDAC multilingual solution through a proof of concept in NCRB. NCRB is evaluating the accuracy and performance of the CDAC's multilingual transliteration engine along with its compatibility for integration with CAS.

In continuation to the above, all States/UTs are advised to undertake activities related to development of business dictionary of police terminology in local language through a committee setup at SCR level, drawing police officers having command over the local language and domain specific knowledge. This would be utilized to improve the accuracy of the multilingual solution in CAS. Kindly refer to the detailed advisory at Annexure 1 in this regard.

Yours sincerely



B Radhika

Joint Director (CCTNS)

23 / 10 / 2012

Copy to:

1. JS(CS), MHA for kind information

Annexure 1:

Development of business dictionary specific to Police Domain for improving the accuracy of CDAC's transliteration engine to be implemented in Multilingual facility in CCTNS

The Core Application Software (CAS) is envisaged to provide the multilingual functionality in following areas:

1. CAS to support user interface and data entry/retrieval in regional language at State level.
2. Data search between Centre and State to have capabilities to transliterate amongst various languages and display the result in desired language (Translation of free running text /unstructured data will not be covered)

Based on the interactions with the various stakeholders at Ministry of Home Affairs (MHA) and during various meetings the requirement of implementing the multilingual capability has been changed from State level to central level. In the TEC (technical evaluation committee) meeting dated 24th February 2011, SDA (software development agency) asked to take over the complete implementation of multilingual capability for all applicable languages into base CAS.

CDAC has been identified as the solution provider for this multilingual solution. Before the complete multilingual solution is implemented in the application, a POC (proof of concept) on the live application is being undertaken at NCRB to ascertain the validity of the solution and the possibility of the implementation of this plug-in with CAS. This POC is being undertaken for two regional languages i.e Bengali and Gujarati.

The accuracy of the transliterated data as accessed in current PoC is less than 70%. This puts a big limitation in terms of usage of transliterated data for business purposes. In order to ensure that the searches and the data entry is faster and accurate, business specific dictionary for police domain needs to be prepared which needs to be transliterated in the regional language which would enable faster and easier transliteration with greater accuracy.

The States/UTs are advised to undertake the following activities to build up the business dictionary specific to police domain:

1. Constitute committee at SCRB level consisting of senior level police officials. This committee would undertake the activities as per the following items:-
 - I. Identify experts from State Police who are well versed in local language as well as in police domain terminology.
 - II. Form the team of above experts which would work to identify various terms used in policing in local language. This team would record these terms in excel sheet along with the English equivalent for each local term.
 - III. The team formed in point II above would also translate the current masters (Available at Annexure 2) in local language.
 - IV. The team formed in point II above will also compile a set of commonly used proper nouns such as accused and convict names.
 - V. The committee would review the progress of activities defined in point II, III and IV on periodical basis.

- VI. The committee would review and approve the outputs generated from activities defined in point II, III and IV.
2. Provide the outputs approved by the committee to NCRB which in turn would be incorporated in the multilingual solution for CAS.
 3. States/UTs need to ensure that this business dictionary is updated on regular basis once the CAS solution is deployed in State/UT. This would ensure that the transliterated results have higher accuracy.
 4. States/UTs need to choose the local language as per the official record keeping language in which they intend to use CAS.
 5. All advanced States in addition to the above, would be required to integrate the multilingual solution in their respective States if current police software does not have data in English.