# Terms of Reference (TOR)

# Background

In this document, we discuss the functional and technical requirements of the Spell and Grammar Checker. Also, we present the general requirements that should be followed for development and release of such a professional standard application. These sections are to be placed as “Scope of Service” of ToR (Section 6 of RFP document).The bidder has to demonstrate her technical understanding and propose methodology for the *technical requirements*. In the “Technical Approach and Methodology” part in Section 5A of RFP, response of the bidder against these requirements should be clearly documented. The deliverables and schedules are presented which should also be placed in ToR. The mode of payments can be used in Section 4 of RFP. The license and copyright issues are also discussed.

We also present a recommended team formulation (which can be extended by the bidder if required) for the research, design, and development of this application and the qualification criteria of members belonging to different positions of the team. This can be used as a guideline in Section 5A. Finally, for the understanding of project management, we present estimated effort of the vendor and overall cost for this application including research and development, supervision, and testing.

# Functional Requirement

* 1. **Spell Checker**
* The spellchecker must be able to correct all types of spelling mistakes such as phonetic spelling mistakes, typos, irregular verb conjugations, similar sound words, misused words correction.
* The spell checker must follow the latest spelling rules declared by Bangla Academy. While doing that, the spell checker must recognize the irregularly formed rules (more commonly known as “নিপাতনে সিদ্ধ নিয়মাবলি”) as well. Alternative valid spellings must be allowed.
* The spellchecker must include correction of (but not be limited to) the following types of errors:

1. Non-word errors. Non-word errors are the errors that result in an invalid word.
2. Real-word errors or contextual spell errors. These errors result in a valid word, which is not used properly in its context. An example is writing “মনে ছিলো আসা” instead of writing “মনে ছিলো আশা”. Here, “আসা” is a valid word, but contextually its use is wrong. Hence, it is a real-word error or a contextual spell error.

* Using the corpus development tool developed by the vendor of another component (SD 7), the vendor of spell and grammar checker has to build an annotated corpus specially designed for this module by injecting large number of errors of all possible types. The corpus must be diverse and representative covering all areas. Common causes of spelling errors have to be identified and studied from relevant literature to generate the errors, e.g., interchange of phonetically similar characters, omission of non-pronounced letters, etc.
* To develop a gold dataset of spelling error, it should be collected from real-world sources and also collected from people of different academic backgrounds and ages.
* In order to obtain sufficiently large error corpus for building spell checker, the vendor has to generate spelling errors automatically developing a supervised learning tool that will use the gold dataset as training data.
* In order to deal with real-word errors, the vendor of spell checker must construct an enriched confusion set of Bangla language. The confusion set is a set of words, which are confusable with the headword of the set because of typing mistakes or phonetic or similarities but are not necessarily confusable with each other.
* Size of corpus should be at leasr10 (ten) million correct and 5 (five) million erroneous (with all required annotations).
* Size of the confusion set should be at least 1(one) million with annotated headwords.
* The corpus must be balanced and representative with inclusion of data from different sources, subjects, domains, etc. Both formal and informal sources should be covered.
* The injected error collection should follow the natural distribution of Bangla language in terms of frequency of words and also expected frequency of error types.
* For correcting errors, separate annotations are required. In order to set up reliable and consistent annotation, several annotators should work independently on the same text. The annotations most agreed upon will be selected.
  1. **Grammar Checker**
* The bidder may assume that the text is free from spelling errors. First, it has to detect unnecessary word, missing word, word or phrase that needs replacement, and words used in the wrong form. To be more specific, the grammar errors to be detected and corrected must include but not limited to the following:

1. Syntactic errors that refer to the sentences or phrases violating the predefined syntax or morphology rules of a language. Example: Subject verb agreement, inappropriate Gender, inappropriate Singular/Plural nouns, Consecutive nouns; Verb tenses relations, prefix-suffix usage, adjective/adverb usage and more.
2. Misspelled words that have been accidentally converted to another valid word.
3. Others like missing punctuation and omission/redundancy of space.
4. Omission of preposition/conjunction, using plurals after cardinal numbers, double plural, repeating word, etc.

* As a reference to different types of Bangla grammar error, we include a list of errors in Appendix A which must be extended substantially in the error tagged corpus.
* Using the annotated corpus development tool developed by the vendor of corresponding component (SD 7), the vendor of grammar checker has to build an annotated corpus specially designed for grammatical error checking by injecting large number of errors of different types. Relevant APIs such as POS tagging, name entity recognition, lemmatization, parsing (both constituency and dependency), tokenization, n-gram etc. will be received from corpus development tool of SD7.
* The vendor has to collect sentences with grammar errors from natural sources such as school going students’ scripts, by giving dictation to people of different literacy level, etc. The errors need to be identified by two linguistic research associates independently and disagreements would be resolved by the expert researcher. Thus a gold set will be developed.
* In order to obtain sufficiently large error corpus for building grammar checker, the vendor has to generate grammar errors automatically developing a supervised learning tool that will use the gold dataset as training data.
* For correcting errors, separate annotations are required. In order to set up reliable and consistent annotation, several annotators should work independently on the same text. The annotations most agreed upon will be selected.
* Both word-level and sentence level error detection and correction need to be performed.
* Size of sentence corpus should be at least 100,000 correct and 5000 erroneous (with all required annotations).
* The corpus must be balanced and representative with inclusion of data from different sources, subjects, domains, etc. Both formal and informal sources should be covered.
* The injected error collection should follow the natural distribution of Bangla language in terms of frequency of words and also expected frequency of error types.

# Input-output specifications

* 1. **Spell Checker**
* Annotated corpus of with automatically injected erroneous words.
* Annotated confusion set that is to be used with SMT for grammar checking.
* Once typing a word is complete, the error(s) should be detected and shown with underline of a visible color.
* The error(s) should be accompanied with a colorful suggestion box at appropriate place around the place of occurrence. There will be user-friendly and standard way to accept/ignore the suggestion.
* Suggestions for correction should be available with “right click” of mouse and the selected option will replace the erroneous word.
* In common cases, the reasons for error should be shown.
* Tool for different platforms, i.e., Windows (32/64 bit OS), Linux, mobile OS such as Android, iOS.
  1. **Grammar Checker**
* Annotated corpus of with injected erroneous sentences and also sentences for error correction.
* Once typing a sentence is complete, the errors should be detected and shown with underline of a visible color.
* Suggestions for correction should be available with “right click” of mouse and the selected option will replace the erroneous text.
* In common cases, the reasons for error should be shown.
* Tool for different platforms, i.e., Windows (32/64 bit OS), Linux, mobile OS such as Android, iOS.

# Technical Requirement

* The component must be compatible with MS office, popular browsers and Desktop Publishing (DTP) software.
* The software should be compatible with major OS (All versions of windows, Linux/Unix, Macintosh, android) and browsers (like Firefox, IE, Chrome, etc.). And also must compatible with major Devices (Mobile, PC, Laptop, Notebook, etc.).
* The vendor needs to generate comprehensive morphological rules for Bangla.
* The standard methods reported in literature such as syntax based, rule-based and Statistical Machine Translation (SMT) based, supervised/unsupervised, feature based/feature independent approaches and their combinations need to be explored addressing all required/possible optimization of techniques.
* The vendor must study contemporary relevant literature, implement the best approaches and use innovation to improve those.
* The vendor must release APIs for different intermediate components that can be tested independently and used by other developers/applications.
* Along with techniques suggested in literature, it is required to develop and implement new innovative techniques to achieve desired level of accuracy.
* The vendor must develop tools and techniques to build separate, representative, and balanced test set that can be used by the external testing team as well to generate new test data.
* Desired performance:

1. Spell Checker, non-word errors:
   * Detection recall: 0.95
   * Correction recall: 0.9
   * Detection precision: 0.95
   * Correction precision: 0.95
2. Spell Checker, real-word errors:
   * Detection recall: 0.8
   * Correction recall: 0.75
   * Detection precision: 0.65
   * Correction precision: 0.75
3. Grammar Checker:
   * Detection recall: 0.7
   * Correction recall: 0.55
   * Detection precision: 0.9
   * Correction precision: 0.8

## Result validation procedure and expected performance

* The vendor has to build testing sets approved by the project team.
* Evaluation of the application need to be measured from average of 100 (or) more experimental round with *k*-fold cross validation.
* The test dataset has to be carefully chosen with sufficient diversity and representativeness.
* Performance has to be evaluated in terms of precision and recall with a target of achieving the desired ones (mentioned before). For comparing different models/classifiers, standard measures such as F-0.5/F-1 may be used.

# General Requirement

The proposed software modules that will be developed in this project will be highly technical in nature and will require extensive research. Thus the development and deployment cycle of these software modules will be different in nature than that of the traditional software application. In this document, we summarize the key technical development guidelines for the project.

## State-of-the-art and scientific methodologies: Most of the components of the software require extensive research and development. Also, selecting appropriate techniques/algorithms is vital to the success of the project. Hence, the vendor should follow the following guidelines:

* The chosen methodologies and techniques should be justified based on the recent high-quality research papers on that domain.
* Every algorithm and methodology should be well documented in a technical paper written by the vendor with proper references to the source articles. The technical document should also describe the rationale behind choosing the appropriate algorithm in right place and describe the competitive advantage(s) over other related algorithms/techniques.
* If any open source module/package is used, the vendor must mention source and the required modification that they need to make the open source module/package suitable for their purpose.

## Platform independent and cloud-ready: All technologies and programming languages used to develop the proposed modules should be platform independent. The developed modules must have the following key features:

* All modules should be deployed on the Amazon/Google (or equivalent) cloud with required sub-systems for the testing period where different external users will test the system.
* The final output should be accessed from all major platforms such as different types of desktop browsers e.g. chrome, IE, Safari, and Mozilla, and different types of mobile platforms such as Android, Windows, Linux/Unix, and iOS. Note that the applications will be hosted in the Data Center of BCC.
* UX design must follow the standard UI design guidelines and adaptable for different screen and device
* All APIs should be accessible for developing new applications in web and mobile.

## Software architecture: The desired characteristics of software architecture are as follows:

## It should be highly modular (independent component) such that similar functionalities are kept in the same sub-module and can be interfaced seamlessly with other sub-modules of the same module or from different modules.

## It should be developed as *n*-tier architecture with different layers for user interface, operational logic, and database (if applicable).

## Components in each layer should be reusable across different modules through APIs.

## Secure design/implementation: The system is expected to address following requirements:

## Development strategies to mitigate known attacks such as cross-site scripting, code injection, buffer overflow, SQL injection by using appropriate framework and handling programmatically should be adopted.

## Appropriate measures will be taken for ensuring security of mobile applications.

## Performance optimization: Different standard performance metrics will be used to test the efficiency and effectiveness of the proposed modules. The developed modules should have the following features.

* The vendor should design the test-cases of each module with different representative dataset to ensure the performance of the module for different metrics.
* The performance of each module will be evaluated both in terms of efficiency (response time or running time) and effectiveness (accuracy of the module).
* All techniques/algorithms/modules must perform much better than the best-known baseline methods according the existing literature.

## Coding convention: The software development has to follow coding convention for the source code and documentation to improve the readability, quality and maintainability of the software following the guidelines given below:

## A well-planned coding and naming convention is important to reduce the effort needed to read and understand the source code, maintain and extend the source code.

## The convention should be designed to keep consistency within a development team, reconcile integration conflicts, and enhance the aesthetic and professional appearance of software.

## Version control and management: The project has to track, make, and manage revisions of the source code in an organized and efficient way.

## Project management and bug tracking: The use of professional project management and bug-tracking tools (such as Confluence and Jira) are recommended for this project. Relevant requirements:

## The vendor should manage all components of the project in one place creating, prioritizing and resolving bugs made easy & traceable by the tool.

## The vendor should organize bugs into components for grouping into smaller parts, and into versions for scheduling activity for milestones or releases. The project monitoring team is expected to be able to browse projects to measure progress and view.

## With Activity streams & Change Log, detailed changes across bugs, issues, projects or people can be followed.

## The vendor should also track important milestones. There has to be provision to assign, track, and complete targeted actions like commenting on page.

1. **Handover of source code and relevant materials:** All source codes, dataset (training and testing) and any relevant material required to reproduce results must be submitted to the client with detail documentation.
2. **Requirement analysis:** The bidder will follow the functional and technical requirements prepared by the consultants from CSE, BUET and receive idea about the scope of the system. Once selected, the vendor will study relevant state-of-the-art literature and identify appropriate methodology to implement solution for the requirements. All their findings from literature need to be documented and selected methodology should be presented before the project team with justification.
3. **Design:** The different steps of design and development of the software must be documented very carefully. This is very important for long term maintenance and sustainability of the system. Based on the approved requirement analysis report, detail system design will be prepared and presented before the consultants for their approval. Note that the applications will be available for both desktop (Windows/Linux) and mobile phones (Android and iOS). User Interface and other design should be done keeping this in mind. It would include, but not limited to the following:

* The following UML and other design components need to be developed at the beginning of the project for every module:
* Use case diagram and description
* Class diagram
* ERD and table attribute description if necessary (SQL Power Architect may be used for drawing ERD)

These will be reviewed and approved by the project team prior to the start of development. Standard tool such as ArgoUML, Dia, Enterprise Architect, or Visual Paradigm should be used.

* All UI prototypes will be designed and presented. The team of consultants will review and approve those.
* Design of interfaces for mobile application must be separately presented.
* All security measures in design and coding should be documented and presented before the start of the development.
* All designs prepared to enhance performance of the system should be documented and presented at the early stage of development.
* Coding convention and documentation stylehas to be designed including purpose of methods, method parameters and flow of code, etc. must be documented in a standard way. Codes should be commented where the comments should explain complex sections of code and provide a rationale for the coding method used. The vendor must submit a sample coding documentation of a method. Use of Doxygen is recommended for this purpose.

1. **Development:** There will be software versions developed for Windows, UNIX, Android, iOS, etc. The vendor has to develop the software complying all the functional requirements approved after Requirement Analysis phase and technical requirements suggested or documented in the *Technical Guideline* to be provided by the project team. The whole system development will be monitored in a very regular and methodical way. Some of the practices to be followed are as follows.

* Version control system will be used as a code repository and for sharing code with consultants.
* The full project will be monitored using appropriate project monitoring and management tools.
* The requirement analysis document and design documents will be reviewed and discussed in joint meetings every week or whenever needed.
* The project manager and relevant team members will need to present the progress of development in every week.
* Code review tools (such as Gerrit) will be used to check if the coding standard is maintained. The source code must be annotated before review begins and it will be verified that review outcomes are addressed properly. The following possibilities will be reviewed:
  + Coding style
  + Project design violation
  + Misunderstood requirements
  + Security defects
  + Inadequate input validation
  + Unsafe methods
  + Lack of exception handling
  + Inefficient programming

1. **Testing:** In this project, we will adopt a test driven development approach. A separate quality assurance team from the vendors will work independently parallel to the development team for this purpose. Project team and the quality assurance team will work together to generate the test cases. However, the quality assurance team will be mainly responsible for performing the tests and submit reports to the project team. Test cases will be written for a number of most important and crucial parts of the system right from the beginning of development. The developers will keep developing the system and try to pass these initial test cases to conform the correctness of the system. The test-cases will be reviewed by the team of consultants as soon as they are written. After development is complete, the full application will be tested comprehensively. The test outcomes of the vendor will be monitored and consultants’ own testing will go on in parallel. Here we discuss different types of testing to be performed in this project.

* UI Testing  
  User interface testing is a testing technique used to identify the presence of defects in a product/software using Graphical user interface [GUI]. There are three basic approaches for UI testing and all of these approaches will be adopted in this project. The approaches are:
* Manual- In this approach the system is tested manually based on the domain and application knowledge of the tester.
* Capture and Replay - In this approach different automation tools are used to test the application in two steps: record and playback. During record, test steps are captured by the automation tool. During playback, the recorded test steps are executed on the system. Tools like Selenium, Watir, Appium (for mobile app) are encouraged to be used for this task.
* Model-based testing - Based on the execution of user sessions based on a GUI model.
* User Acceptance Testing

Since these tools will be used by the mass people, the vendor must test it by ordinary people considering it as mock user acceptance testing. The selected users for testing should be diverse in consideration of their knowledge and maturity and at least 10 in number.

* API and Performance testing

API Testing will involve testing application programming interfaces (APIs) directly and as part of integration testing to determine if they meet expectations for functionality, reliability, performance, and security. API testing will be considered critical for testing in this project because APIs serve as the primary interface to application logic. The vendor will perform extensive API tests to confirm the correctness of the functionalities and identifying code segments/ function creating bottlenecks in terms of performance. The APIs that may be used by other modules in this project or in future must be tested very carefully.

* Security Testing

Since the vendor will need to deploy the tool along with help instructions under subdomain of a common domain, they have to ensure the server is not compromised through their page. For mobile apps, any chance of security breach such as accessing other resources must be mitigated.

In order to ensure the security of the system, it is necessary to design a security development lifecycle. Security should be considered and tested throughout the project lifecycle. Security testing will be performed to ensure that the system protects the data and maintains its intended functionality. The security testing will involve an active analysis of the application for any weaknesses, technical flaws, or vulnerabilities.  The primary purpose will be to identify the vulnerabilities, and subsequently repairing them. The vendor team will perform extensive tests to check the system for security vulnerabilities and critical portions of the source codes should also be reviewed for known security threats. The tests carried out by the vendor team will include from the following security vulnerabilities wherever applicable:

* SQL Injection
* Cross Site Scripting
* Broken Authentication and Session Management
* Insecure Direct Object References
* Cross Site Request Forgery
* Invalidated Redirects and Forwards
* Sensitive Data Exposure
* Using components with known vulnerabilities

There are several tools available for security testing e.g., Zed Attack Proxy (ZAP), Metasploit, Vega, SQLMap which should be used for testing the application for security vulnerabilities. The vendor team is required to provide a brief plan of their security testing. The plan should explicitly list the vulnerabilities that the vendor plans to address and the tools they will be using to detect them. The plan should also contain a separate part specific to mobile devices (Android and iOS).

1. **Using open source database, e.g.,PostgreSQL, MariaDB, SQLite:**On principle, if any Database Management System (DBMS) is required, it must be free/open source. PostgreSQL, the open-source database platform quite popular with web and mobile application developers for its ease of maintenance, cost effectiveness, and simple integration with other open-source technologies, is recommended for this project. The vendor has to provide own-built/free utility to back up its database regularly. The system must be user friendly ensuring the following:

* Safeguarding against data loss due to failure of underlying infrastructure components like storage or the server itself.
* Safeguarding against data corruption and unwanted or malicious loss of data.
* Migrating production databases into development or test environments.

The vendor has to install two PostgreSQL servers in two different machines. One of the PostgreSQL servers will be our main database server where the vendor will create production database. The second PostgreSQL instance will be empty and treated as a standby machine where the vendor can restore from the backup. For taking Backup, the vendor has to install a third server and make the whole system automated.

1. **Maintenance:** The vendor has to maintain and update the software after its delivery to the client in order to correct faults and to improve performance or adapt to a changed environment. Relevant requirements:

* For any updated version, automatic notification should be sent to all registered users with provision of easy and secure update.
* There has to be a dedicated skilled maintenance team capable of updating the software to fix bugs discovered by the maintenance team or the users in a very short time.
* All the changes made to the software should be properly logged and approved by a designated approving authority (e.g., the head of the maintenance team).
* The maintenance team must respond to BCC project team immediately if a bug is found.
* Total maintenance period is 2 years and 6 months. However, maintenance of the first six months is free of cost.

1. **Technical Report:** The vendor must submit a technical report describing the methodology, related work, experimental design, etc. All steps or components must be discussed in a way that future researchers would be able to use those.
2. **Training/transfer of knowledge:** Since these applications will be used mostly by the mass people, very well designed user manual for installation and use should be provided in both Bangla and English. Also, these will be open source products to be used/interfaced by other applications in future. Separate manuals need to be developed to facilitate this purpose. Detail requirements are as follows:

* User manuals for general users need to be developed by the vendor for effective use of features and installation in all available platforms, e.g., Windows, UNIX, Android, iOS, etc. The manual should also cover application version upgradation. In some cases, to explain complex operations, video tutorials may be suggested by the project team.
* Separate training manuals need to be developed for administrative users of the project maintenance team who may augment corpus and make enhancements.
* Many applications may need to be interfaced with the application. For this purpose, API documentation and technical documentation need to be developed maintaining professional standard.
* After the completion of the maintenance period with the vendor, a team of in-house software developers may be appointed to enhance the software and accommodate change requests. This team should be trained about the software architecture, design, coding convention, etc. so that future enhancement and maintenance may be done by them.

# Deliverables and Schedule

The deliverable tasks and expected schedules are given below considering **18 months** of development period:

|  |  |  |
| --- | --- | --- |
| **Task** | **Desired size/ performance** | **Time of delivery (month)** |
| Development of tool for automatically generating error injected corpus and the corpus in desired size. | Spell Checker: 10 million correct and 5 million incorrect words  Grammar checker: 100,000 correct and 5,000 erroneous sentences | 4th |
| Development of tool for generating corpus for correction of error and the corpus in desired size. |  | 4th |
| Implementation of spell and grammar checker’s first version. | Spell checker, non-word errors:   * Detection recall: 0.85 * Correction recall: 0.75 * Detection precision: 0.85 * Correction precision: 0.8   Spell checker, real-word errors:   * Detection recall: 0.65 * Correction recall: 0.6 * Detection precision: 0.6 * Correction precision: 0.65   Grammar checker:   * Detection recall: 0.55 * Correction recall: 0.4 * Detection precision: 0.75 * Correction precision: 0.65 | 8th |
| Implementation of spell and grammar checker’s final version. | Spell checker, non-word errors:   * Detection recall: 0.95 * Correction recall: 0.9 * Detection precision: 0.95 * Correction precision: 0.95   Spell checker, real-word errors:   * Detection recall: 0.8 * Correction recall: 0.75 * Detection precision: 0.65 * Correction precision: 0.75   Grammar checker:   * Detection recall: 0.7 * Correction recall: 0.55 * Detection precision: 0.9 * Correction precision: 0.8 | 14th |
| Integration of the model with MS office/open document format, popular browsers and DTP software |  | 16th |
| Implementation of versions for other than the first platform (OS) |  | 17th |
| Mass-level testing after public release and reception of all documentations |  | 18th |

# Payment Method/Phase

Recommended payment against deliverables:

|  |  |
| --- | --- |
| **Task** | **Payment** |
| Development of tools for automatically generating error injected corpus and corpus for correction of error. Also, both the corpus in desired size: upon submission of the Inception Report | 5% |
| Implementation of spell checker’s first version. | 10% |
| Implementation of spell checker’s final version. | 10% |
| Implementation of grammar checker’s first version. | 10% |
| Implementation of grammar checker’s final version. | 15% |
| Integration of the model with MS office, popular browsers and DTP software, implementation of versions for other than the first platform (OS), and mass-level testing after public release and reception of all documentations. | 50% |
| 2 years maintenance | Additional 12% |

# License and Copyright

Ownership of the deliverables and all related matters will be under the following definitions and clauses:

## 8.1 Definitions:

* "Intellectual Property" (hereinafter referred to as "IP") shall mean to include but not limited to, all software/reports, designs, programs, specifications, documentations, manuals, visual aids, patent rights, copyrights, database rights, trademarks, service marks, business names, Internet domain names, registered designs, design rights, topography rights, mask works, and any other similar protected rights in any country together with pending applications or capable of being registered or recording of such rights.
* "Project IP" shall mean the Intellectual Property that is conceived (in the case of inventions) or created (in the case of Intellectual Property other than inventions) in the performance of the assigned Project.
* “Background IP” means IP independently conceived or reduced to practice or acquired by the vendor before execution of this Agreement used for Foreground IP.
* “Foreground IP” means any and all inventions, devices, processes (including but not limited to processes of using devices or of manufacturing such devices), methods, utilities, compositions or products or software, whether patentable or unpatentable, and works of authorship, and related know-how, which may be acquired or developed during the term of this Agreement and for one hundred eighty (180) days after this Agreement expires or terminates, as a result of conducting Project by the vendor under this Agreement.
* No third party branding is allowed. The tools will contain only the logos of the Government of Bangladesh, ICT Division, BCC, and the project.

## 8.2 Clauses:

* All Background IPs shall be transferred by the vendor to the purchaser, and will be regarded as the sole and exclusive property of purchaser after commencement of the project. Since then, no license on the Background IP could be issued by the vendor. All the in-operation or pending licenses on the Background IP already bestowed by the vendor must be retracted. The vendor will be solely liable for performing the retraction tasks. Any financial involvement, which could be associated with the purchaser, for such retraction tasks shall be borne solely by the vendor.
* The vendor shall in good faith notify the purchaser about the vendor’s all Background IPs that the vendor is aware of.
* The vendor shall promptly furnish the purchaser with an invention disclosure(s) of any Foreground IP. Only the purchaser may, at its sole discretion and expense, initiate and control the preparation, filing, and/or prosecution of patent applications or other protective measures concerning Foreground IP in any country with or without any prior notification to the other.
* In case the vendor develops inventions that are related to Project but are not covered by the scope of Project, the vendor will grant an exclusive option to the purchaser to license exclusively such inventions and corresponding or future patent rights thereon. The Parties will negotiate in good faith a corresponding license agreement.
* The purchaser will be the sole owner of all the Foreground IPs that shall come up from this project. Any beneficial interest, patent filing, commercialization, or similar tasks will be done at the sole authority of the purchaser. The vendor cannot use the Foreground IPs in any of these tasks or similar.
* No second/third party branding is allowed. The tools will contain only the logos of the Government of Bangladesh, ICT Division, BCC, and the project.
* Consultant/s has to use the Icon/Logo of the component provided by the Project/BCC/ICT Division.

# Required qualification of team members

While selecting members of the project team dedicated to carry out this assignment, the bidder should consider the following guidelines.

|  |  |  |  |
| --- | --- | --- | --- |
| **Position** | **Minimum required person** | **Educational qualification** | **Experience requirement** |
| Team leader/manager | 1 | CSE/CS graduate from reputed university (PhD in the relevant field would be a plus point). | 10+ years’ experience in technical and senior managerial role. Experience in development of language processing tool will be an added advantage. |
| Linguistic expert/researcher | 2 (may be part-time consultant) | Masters/PhD in relevant subjects with quality research. Special interest and experience in grammar is preferred. | * Experience with computational linguistics. Experience to design/use computational linguistic tool is highly preferred. * Relevant and high-quality research publications must be mentioned in CV. Note that the relevant strength of the department will be considered during evaluation. |
| Linguistic associate  (To work under guidance of expert for generating annotated dataset (i.e., error tags in corpus ) | 4 (2 for grammar and 2 for spell checker) | M. Sc. in relevant subject. |  |
| Computer Science Researcher with research experience in Machine Learning, Language Processing, or Applied Machine Learning fields. | 2 (may be part-time) | PhD in CS/CSE or equivalent area from any reputed university with research experience in Machine Learning, Language Processing, or Applied Machine Learning. | * Several research publications in Tier 1/ Tier 2 journal/conference proceedings according to acceptable ranking such as CORE 2017/ ERA 2010 (to be presented in CV) are preferable. * Experience with design/development of language processing or applied machine learning based software. * Experience in software design/development. |
| Lead developer/Technical lead/software architect | 1 | CSE/CS graduate from any reputed university. | * 7+ years’ experience in technical and team-lead role. * Experience in relevant project is considered added qualification. |
| Senior Software Engineer | 4 | CSE/CS graduate from any reputed university (two of them should have hands on experience in project development using machine learning / AI techniques, one of them should have experience in Amazon cloud architecture and deployment of software in the cloud). | * 5+ years’ experience. * Experience in relevant project/research is considered added qualification. |
| Software Engineer | 6 | CSE/CS graduate from any reputed university. | 1+ years’ experience (for relevant research experience from very reputed university, experience requirement is relaxed.) |
| Tester | 2 | CSE/CS graduate from any recognized university. | Should have 3+ years’ experience |
| UI designer | 1 | Bachelors from relevant subject. | 5+ years’ experience. |
| Documentation staff | 2 | Bachelors from relevant subject with high language proficiency. | 2+ years’ experience. |
| Support associates (to facilitate data collection such as grammar error example and processing those) | 4 | B.Sc./diploma in IT related subjects. | 1 year experience |

\*\* Notes:

* All positions mentioned above are full-time. However, the vendor can appoint equivalent number of part-time personnel if necessary with the consultation with the project authority.
* For extra-ordinary experience and relevant achievement (e.g., success in ACM programming contest), team members having non-CSE/CS degrees will be considered.
* Also, for extra-ordinary qualification, experience requirement may be relaxed.
* Bidders may add more personnel and/or more positions if they think necessary.
* Educational qualification can be relaxed if anyone has exceptional knowledge and experience in any domain.

# Cost and Effort Estimation

**Development period:** 18 months.

## Effort Estimation of the Vendor for Research and Development

|  |  |  |  |
| --- | --- | --- | --- |
| **Position** | **Minimum required person** | **Duration of involvement (month)** | **Estimated man-month** |
| Project leader/manager | 1 | 18 | 18 |
| Linguistic expert/researcher | 2 (may be part-time consultant) | 18 | 18 |
| Linguistic associate  (To work under guidance of expert for generating annotated dataset (i.e., error tags in corpus ) | 4 (2 for grammar and 2 for spell checker) | 14 | 56 |
| **Computer Science Researcher with research experience in Machine Learning, Language Processing, or Applied Machine Learning fields.** | 2 (may be part-time) | 18 | 18 |
| Lead developer/Technical lead/software architect | 1 | 18 | 18 |
| Senior Software Engineer | 4 | 18 | 72 |
| Software Engineer | 6 | 18 | 108 |
| Tester | 3 | 12 | 36 |
| UI designer | 1 | 4 | 4 |
| Documentation staff | 2 | 6 | 12 |
| Support associate | 4 | 6 | 24 |
| Total | | | 384 |

## Effort Estimation for Testing by External Experts

The external experts’ effort for testing at different phases including test-case generation and test dataset development is discussed below.

|  |  |
| --- | --- |
| **Test Item** | **Estimated man-month** |
| Study of relevant literature and make detail test plan | 1 |
| Testing the balance and representativeness of error injected corpus of desired size for the spell and grammar checker. | 1 |
| Testing the tool for generating corpus for correction of error and the size/quality of corpus for the spell and grammar checker | 1 |
| Testing spell and grammar checker’s first version with tester generated representative dataset. | 2 |
| Testing spell and grammar checker’s final version for first OS with tester generated representative dataset. | 1 |
| Testing Integration of the model with MS office/open document format, popular browsers and DTP software. | 1 |
| Testing versions for other than the first platform (OS) and receiving final deliverables (code, dataset, documentation) | 1 |
| Total | 8 |