Manual Search within the Document impacts overall efficiency of operations of MoSPI

Details	
Particulars	**Details**
Name of Ministry/Department/Division	Ministry of Statistics and Programme Implementation/Computer Centre
Address	10, East Block, Rama Krishna Puram, New Delhi, 110066
Name of the Nodal Officer and Designation	Ms. Madhura Roy, Director
Phone Number (Nodal Officer)	8860570648
Email ID	madhura.roy@gov.in
Domain/Area: Problem Statement	Official Statistics in general
Category of problem statement *(Select all that are applicable)*	☐ Data Collection ☑ Data Processing ☑ Data Quality ☐ Statistical Methodology ☑ Data Accessibility ☐ Data Integration and Interoperability ☑ Data Timeliness ☐ Data Standardization ☑ Data Utilisation and Analysis ☐ Data Visualization ☐ Data Transparency ☐ Technological Infrastructure ☑ Data Management ☐ Data Extraction and Pipeline Creation ☐ Others (please specify)
What kind of support do you expect from the solution giver?	☐ Research and development of new methodology ☑ Development or modification of tools/process
What kind of support/resources will you be able to share with MoSPI?	☑ Subject Matter Experts (SMEs) ☑ Existing Datasets ☐ Technical Resources ☐ Collaborative Networks ☐ IT Infrastructure ☑ Physical Infrastructure (space for partners to work from) ☐ Others (please specify)

*Mandatory

For non-mandatory questions, enter "N/A"in the given box if not applicable

Problem Statement

A. Problem statement Identified (Max 200 words) *

(Write a crisp and specific problem statement identified by the Ministry/Department/Division. Include aspects such as what the problem is, whom it impacts, and scale of impact. Try to give data figures wherever possible. Make sure the problem statement is related to official statistics areas.)

Ministry of Statistics and PI during its various activities needs to refer to manuals and other relevant documents mostly available in Portable Document Format (PDFs) or images. Though pdf is searchable but manual document retrieval requires specialised knowledge and consumes considerable time. Even if the manual is available in pdf or any other machine-readable format, search result may not yield the most relevant match and considerable effort is required to find the most pertinent information which remain elusive. For example, investigators often spend substantial time manually searching for specific information, interpretations, and guidelines within a manual during fieldwork, despite comprehensive training provided to them. This may result in productivity losses and delays in data collection and quality, hindering the effectiveness of the data-producing activities. It would also be helpful if there were an AI based tool available which facilitates quick search from the given document based on the input provided in writing or through a voice command. The tool should be able to search both the web or the given document(s).

Contains 171 words

B. Methodology used for identifying the problem statement. *

(Detail how the problem was identified, including any studies conducted, resources referred to, or methodologies applied.)

The problem was identified through a detailed review of existing processes and feedback received from various stakeholders, which highlighted the inefficiencies in manually searching documents. Feedback from officials, staff, and field officers was gathered through informal communication to assess the time taken for keyword searches in PDF documents. Insights from this exercise, combined with an analysis of common challenges faced in retrieving document content from the web, led to the identification of this problem.

*Mandatory

For non-mandatory questions, enter "N/A"in the given box if not applicable

C. Challenges imposed and need for solving them. *

(Explanation of the current situation, including relevant data and statistics that highlights the need for addressing this problem. List all key stakeholders affected by this problem, including internal teams, external partners, or end-users. Highlight the potential long-term impacts if the problem remains unsolved.)

The instructions manuals are typically stored as PDFs, making it difficult for users to efficiently search for specific content and interpretations, which should ideally come up in a handy way. The current process is highly time consuming, requiring individuals to perform keyword searches, and review the results—tasks that are time-consuming and prone to errors.

The lack of a streamlined, automated search process leads to significant delays in finding relevant information, reducing the overall productivity. The reliance on domain expertise means that newer staff or those unfamiliar with specific survey content face additional challenges, further slowing down operations. If left unaddressed, this problem could result in long-term inefficiencies, increased operational costs, and frustration among staff, ultimately impacting the quality of surveys and data driven insights.

Key Stakeholders Impacted:
Various Ministries, Departments, and Divisions including MoSPI, policy makers, partners awaiting crucial information, and end-users who rely on timely decisions, are all adversely affected by this issue.

D. Existing processes/systems in place to deal with the challenges (Max 150 words). *

(How is the Ministry/Department/Division currently addressing the problem statement? In case no way has been found to manage it, kindly mention that as well.)

The document is either available on the web or on the machine used for data collection in PDF/image. The user runs a manual search and uses its knowledge to reach the most relevant information. Overall, **no comprehensive, unified system is in place** to address the problem, leading to inconsistent search experiences and ongoing inefficiencies.

Contains 54 words

*Mandatory

For non-mandatory questions, enter "N/A"in the given box if not applicable

E. Expected outcome(s) for stakeholders' post resolution. *

(Clearly outline the benefits and improvements the impacted stakeholders will experience once the problem is resolved. Also mention the essential features of the solution.)

- **Improved Efficiency:** Quick retrieval of specific content and interpretations through advanced keyword and context-based search will significantly reduce document search time, facilitating faster understanding, decision-making, and easier adaptation to updated manual content.
- **Enhanced User Experience:** User-friendly interface making it easy for field officers to quickly access and become well-versed with manual content.
- **Better Resource Allocation:** Staff and field officers can focus on more critical tasks instead of manual searches.
- **Increased Accuracy:** Fewer errors in retrieving relevant information.

F. How urgent do you consider it to solve this problem? *

(What degree of impact does the problem have on operations?)

☑ **High Priority:** The problem significantly impacts daily operations; needs immediate attention.

☐ **Medium Priority:** The problem affects productivity or efficiency but does not halt operations. It should be addressed within a reasonable timeframe.

☐ **Low Priority:** The problem has minimal impact on overall operations and can be resolved at a later time without major consequences.

G. Expected timeline for resolution?

(Mention the duration within which you expect a resolution.)

A resolution is expected within 6 to 9 months.

H. Share any global best practices you'd like to highlight?

(Mention any global best practices you know of that could address your issue or be implemented to solve it. Include links where possible.)

GPT and other foundation models provides for such functionality through their chat interfaces

*Mandatory

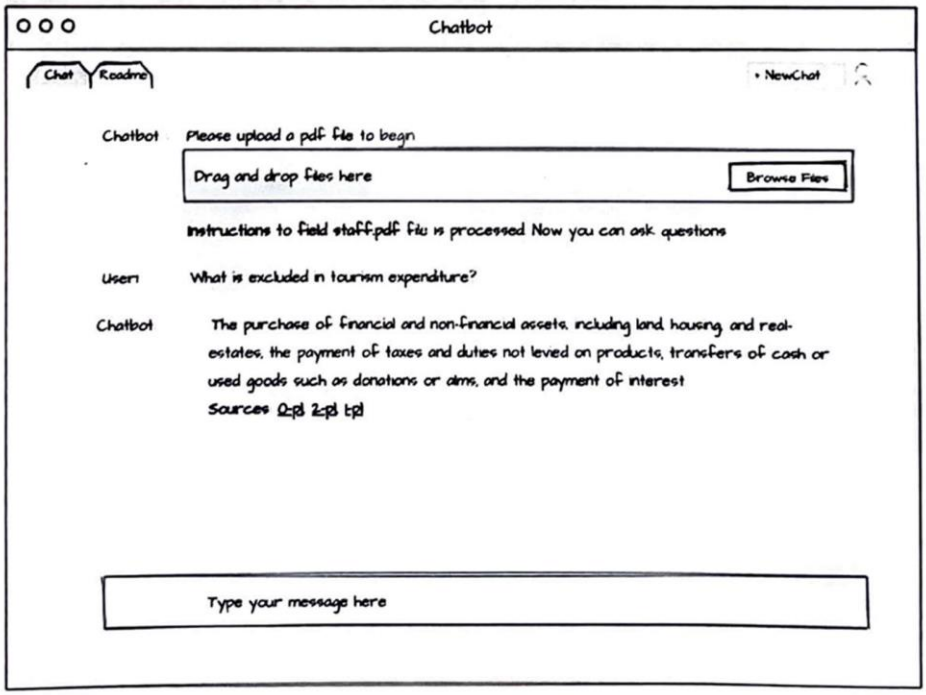
For non-mandatory questions, enter "N/A"in the given box if not applicable

(Complete this section only if the Ministry/Department/Division submitting the proposal has potential solutions in mind)

I. Proposed solutions (Max 200 Words)

(Provide an overview of the proposed solutions, including the key milestones and tentative timelines for each phase of implementation. Try to post your idea in points /diagrams /infographics /pictures.)

The solution is a web-based chatbot that uses AI-powered search based on NLP, users to upload documents or search the web and receive answers to their questions in natural language, whether asked through written or verbal commands. The chatbot should quickly return accurate results, displaying the relevant page from the document to the user. The interface will allow easy document uploading and searching, similar to a provided design, though minor differences are acceptable as long as it functions as required. The bot should be able to use website of the ministry or web in general if required.



J. Analysis of the feasibility of the solution

(Evaluate the viability of the proposed solution, considering its technical, financial, and operational aspects, along with identifying potential challenges and risks.)

The proposed solution is technically feasible, leveraging existing NLP and AI technologies that have proven effective in similar contexts, thereby offering enhanced search capabilities. Operationally, the solution can potentially be integrated with current document management systems/websites, minimizing disruption during implementation.

*Mandatory

For non-mandatory questions, enter "N/A"in the given box if not applicable