



# Text Scrutiny And Visualization Using NLP

<sup>1</sup>Vedant Pachpute, <sup>2</sup>Yadnyali Behere, <sup>3</sup>Mrunali Pagare, <sup>4</sup>Pratik Bhagat, <sup>5</sup>Prof. Vikas Nandgaonkar

<sup>1</sup>Student, <sup>2</sup>Student, <sup>3</sup>Student, <sup>4</sup>Student, <sup>5</sup>Professor

<sup>1,2,3,4,5</sup>Department of Computer Engineering,

<sup>1,2,3,4,5</sup>Indira College of Engineering and Management, Parandwadi, Pune, India

**Abstract:** The market is flooded with text mining and visualization tools that make it easier for creative individuals to find "hidden nuggets" of knowledge regarding developing technology. Every year, more and more data are produced, necessitating the need to synthesize and gather knowledge from an ever-expanding body of literature. In the corporate world, time is more valuable than money, so it takes a lot of time to comprehend what the data actually is, why it was developed, and why it is important. In this paper, we describe a case study of text analytics and data visualization using computational techniques like Natural Language Processing (NLP). This data can occasionally be found in an unstructured text format, which falls under the category of big data and has to be analyzed to yield useful information. This project uses Python libraries to process a large volume of text and provide graphical visualizations and text analysis operations such as Word cloud, Mendenhall Curve, Tokenization, Graph, Processed Text, and Name Entity Recognition (NER). The project shows that text analytics based on NLP is an effective method for comprehending the Data. To save time wastage and effort, this paper makes an effort to give a complete assessment of the data and to make appropriate data-driven judgement. This method allows the user to know the data and revise grammar simultaneously, which may be valuable for meetings, analysts, employees, students, and teachers.

**Index Terms-**Text mining, text analytics, visualization, NLP.

## I. INTRODUCTION

Any organization depends on its data. It serves as the basis for your decisions and improves in your ability to outperform your adversaries. Regrettably, it's also one of the most challenging things to monitor because it typically takes place in back-end databases or spreadsheets that are challenging to access or understand. Natural Language Processing can help with that as well. NLP lets you scrutinize text using computer algorithms to extract information that would be otherwise not possible to comprehend on your own.

Human language study and comprehension is a technique known as natural language processing (NLP). It enables you to draw significant conclusions from vast amounts of data, which can aid in the improvement of business decisions. The process of gaining meaning from text is called text analysis or scrutiny. NLP can help you comprehend your audience's requirements and desires so that you can provide them the knowledge they require to make wise decisions. To gain a deeper understanding of the data and utilize the text in a business setting, text analysis might be helpful.

Text visualization is the technique of illustrating to readers how their data appears in a creative manner using visuals or images. They will be able to better grasp the data and use it to make decisions as a result of their improved ability to visualize it. Text visualization uses data to produce pictures that convey information in order to make sense of what's happening and highlight key details. We simplify the process of making wise business decisions for you by offering data insights and context for important trends. Additionally, we utilized our NLP-based product to maximize productivity and save time wasting. Our product will analyze your data after you specify what you want it to perform. For you to make wise selections, it will quickly but thoroughly uncover the most crucial components. This technology, which is supported by an intelligence platform, contributes in helping you comprehend difficult data and come to better business selections.

## II. MOTIVATION

This project aims to investigate advance NLP research. It is challenging to quickly abstract the data from large and unstructured data, so MNCs benefit from this project's quick understanding of the data. The reader typically reads the information rapidly when there is a time constraint and an important meeting is about to begin, and if it is not prepared swiftly, they are unable to match the points and understand the relationships between words. Text analysis and visualization can make it simple, rapid, and scalable for the user to understand qualitative data. Word cloud, which uses charts, will transform plaintext into captivating narrative. The Text Analyzer is the ideal tool for individuals since it not only summaries the data but also highlights the Nouns, Verbs, and Pronouns in the text, enhancing both readability and learnability.

### III. PROBLEM DEFINITION

Understanding and analyzing textual data is one of the biggest challenges. Every day, there is a significant amount of data generated, and MNCs must work under very tight deadlines with impromptu meetings. Employees don't have enough time to thoroughly read all the data, which has an indirect impact on decision-making. The same situation can be observed in schools and universities, where students are unable to comprehend the teacher's data and notebook because the human brain is not capable of grasping abstract concepts as quickly as graphs and visually represented images.

The human brain's inability to keep information for an extended period of time without summarizing it is another problem. For instance, in countries like India where English is not widely spoken, individuals tend to forget the basics of grammar, such as nouns, verbs, pronouns, and so forth. In order to address the aforementioned issues, we built this project that enables users to quickly understand the data and simultaneously offers both learnability and readability.

### IV. PROJECT SCOPE

1. Companies no longer need to have a pre-requisite detailed theory of data in order to respond quickly to decisions based on data.
2. Employees can use this to highlight key data segments, present data as a wordcloud and other visual formats during stakeholder meetings and assist in data comprehension.
3. By using a visual format and minimal effort, this approach can be used in colleges and universities to teach students challenging concepts.
4. While reading, students can easily comprehend the content and refine their English language skills.

### V. EXISTING SYSTEM

While there are a few NLP tools available in the market, none of them are well-made or equipped with all necessary features. These features can be supported by a wide variety of websites and mobile applications. Since many systems only have one or two features, it is difficult for the user to access each website for various purposes and this takes a lot of time. Unwanted advertisements are the main barrier to using these systems; since the developers don't make any money from them, they must be displayed. Restricted input is another measure that this system is limited by; users are not permitted to upload files in many systems and must copy and paste the data within a set amount of word limit, this is a headache for users. Since numerous systems provide various features, none of them are unified under a single umbrella. Some mobile applications allow users to scan text using their camera, but the results are sometimes unreliable and the data is only partially accurate.

### VI. RELATED WORK

In the area of natural language processing, there has been significant advancement. People are taking data-driven decisions more swiftly thanks to the tools and implementation in this area. There are certain tried-and-true projects being carried out by some organizations in this enormous and occasionally complex world of NLP, but none of them have been properly developed and designed. The project should be built taking into account all the features and giving the user everything under one roof in order to benefit the user rather than confusing them. In this report, we'll discuss how we're creating a fascinating website with cutting-edge NLP techniques. Sublime and the Streamlit framework are being used to construct this project in Python. We will integrate all the crucial functions under one hood, from text analysis to visualization, from name entity identification to part of speech tagger. The reason we chose this project is that there is a great deal of distinctions in NLP tools and their output, making it challenging for users to use various programmes for various jobs and wasting a great deal of time and resources.

### VII. PROPOSED SYSTEM

The Text Scrutiny and Visualization system is a collection of all the cutting-edge tools that enables users to gain insights in just a few clicks, as was previously indicated. The system is split into two phases: the first is input, and the second is applying operations on the data. In the first stage, we have given users the ability to directly upload any file up to 200 megabytes in size, whether it be a text or PDF file. This is a crucial component of the project since it enables users to gain full insights into data in a single step as opposed to breaking it down into smaller modules and copying and pasting it because most systems have word limits that lead to inaccurate findings and also result in time wastage.

In the second stage, the user must apply operations to the data, such as wordcloud, mendelhall curve, Part of Speech tagger (POS), Name Entity Recognition (NER), tokenization, text analysis, graph, and text summarizer.

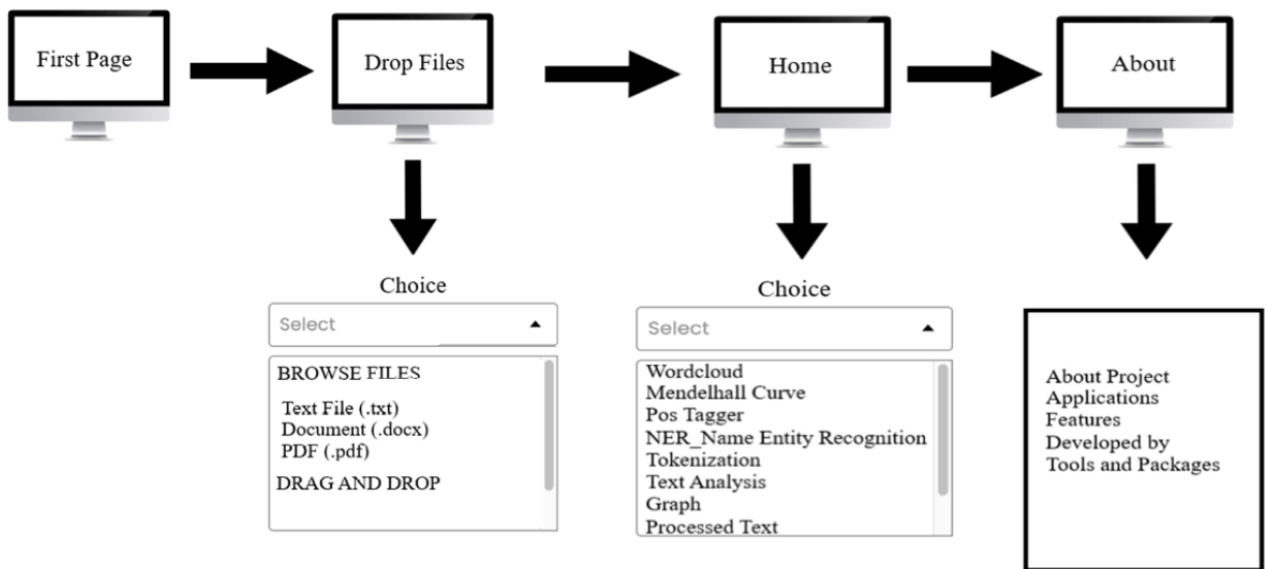


Figure 1: System Architecture

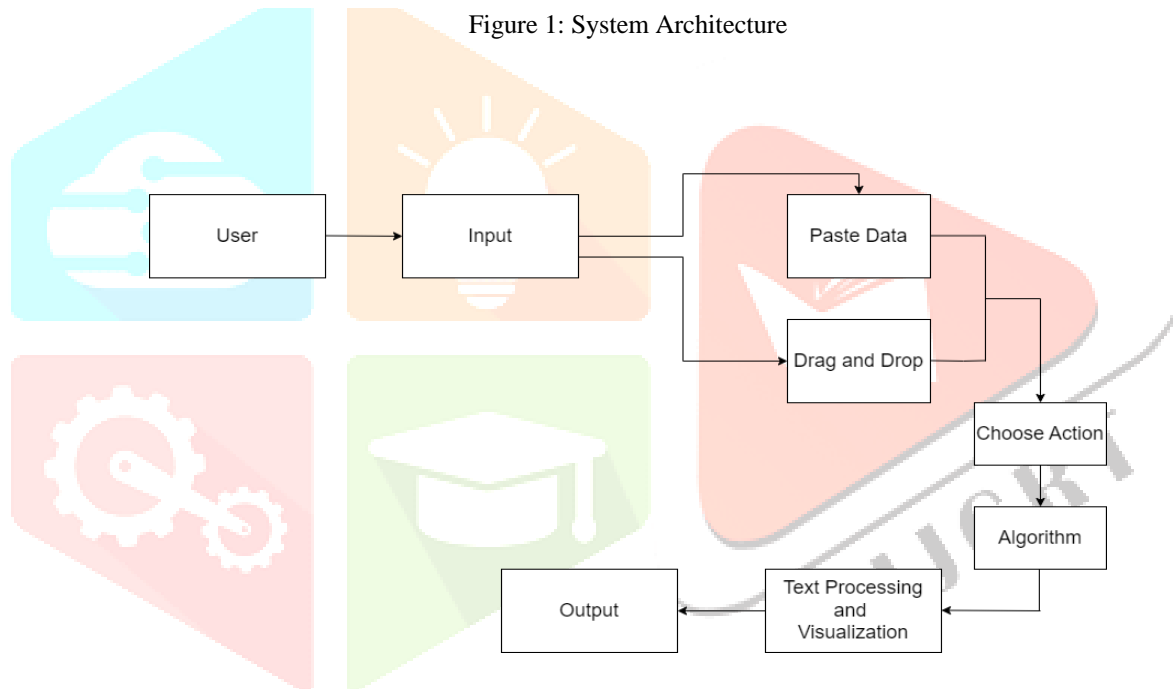


Figure 2: Working Methodology

VIII. SNAPSHOTS

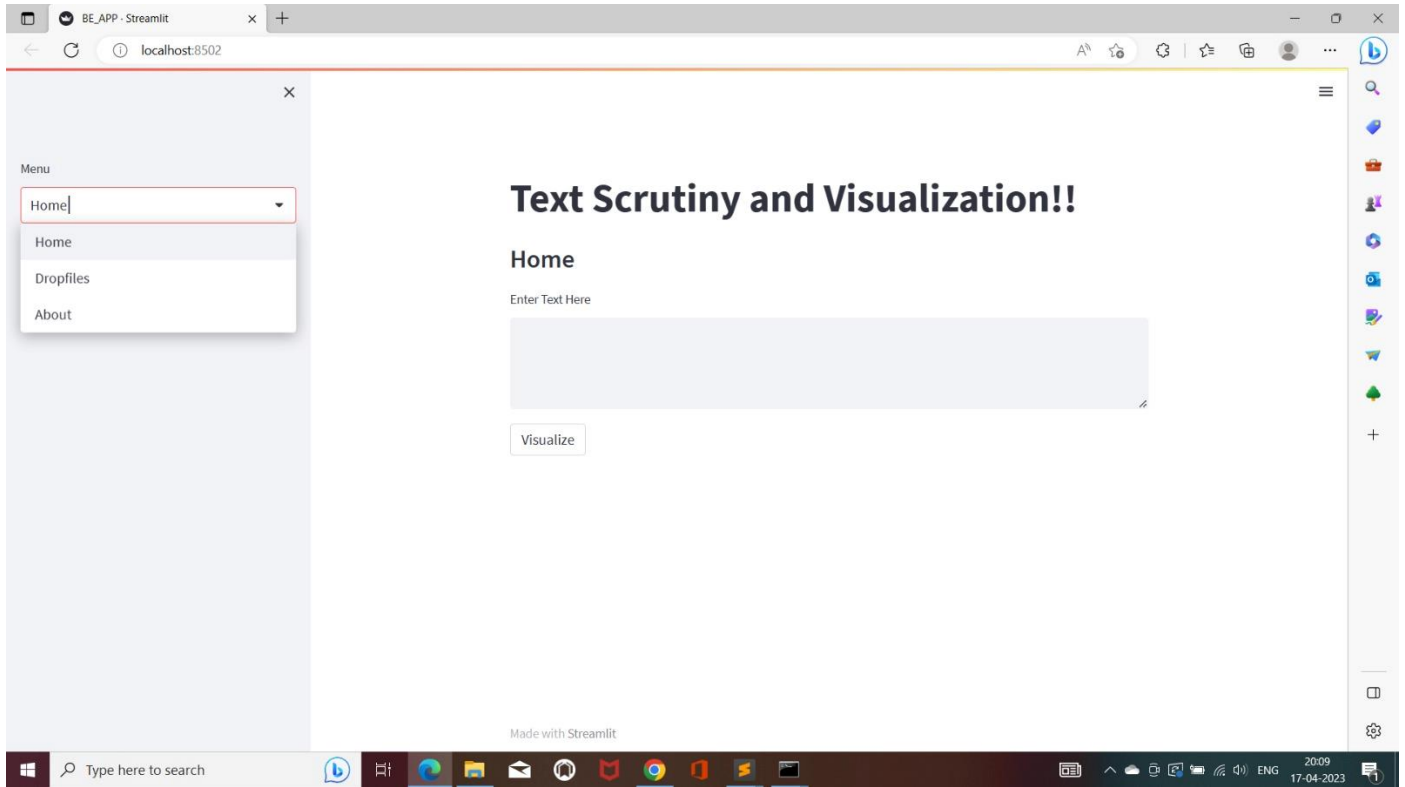


Figure 3: Home Page

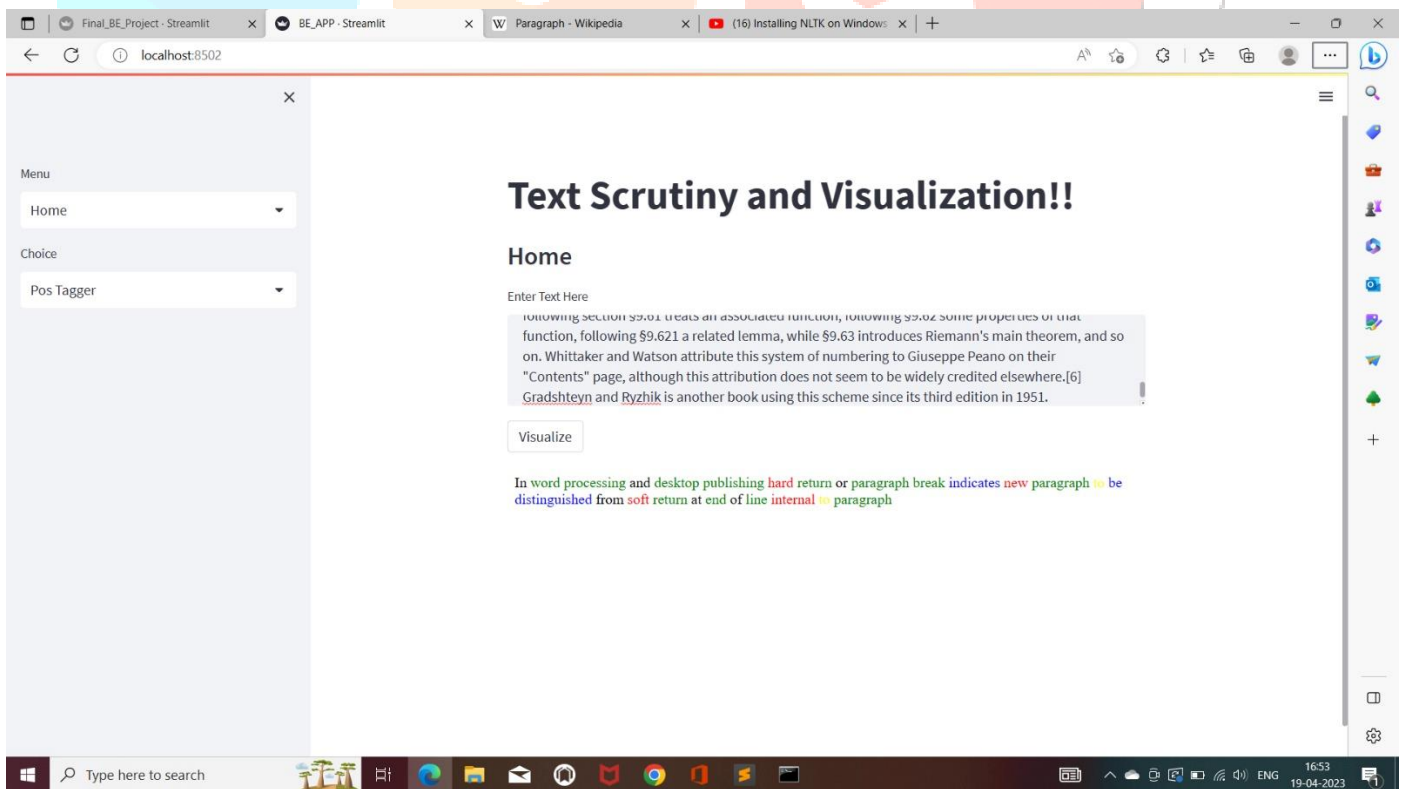


Figure 4: Demonstration of POS Tagger function

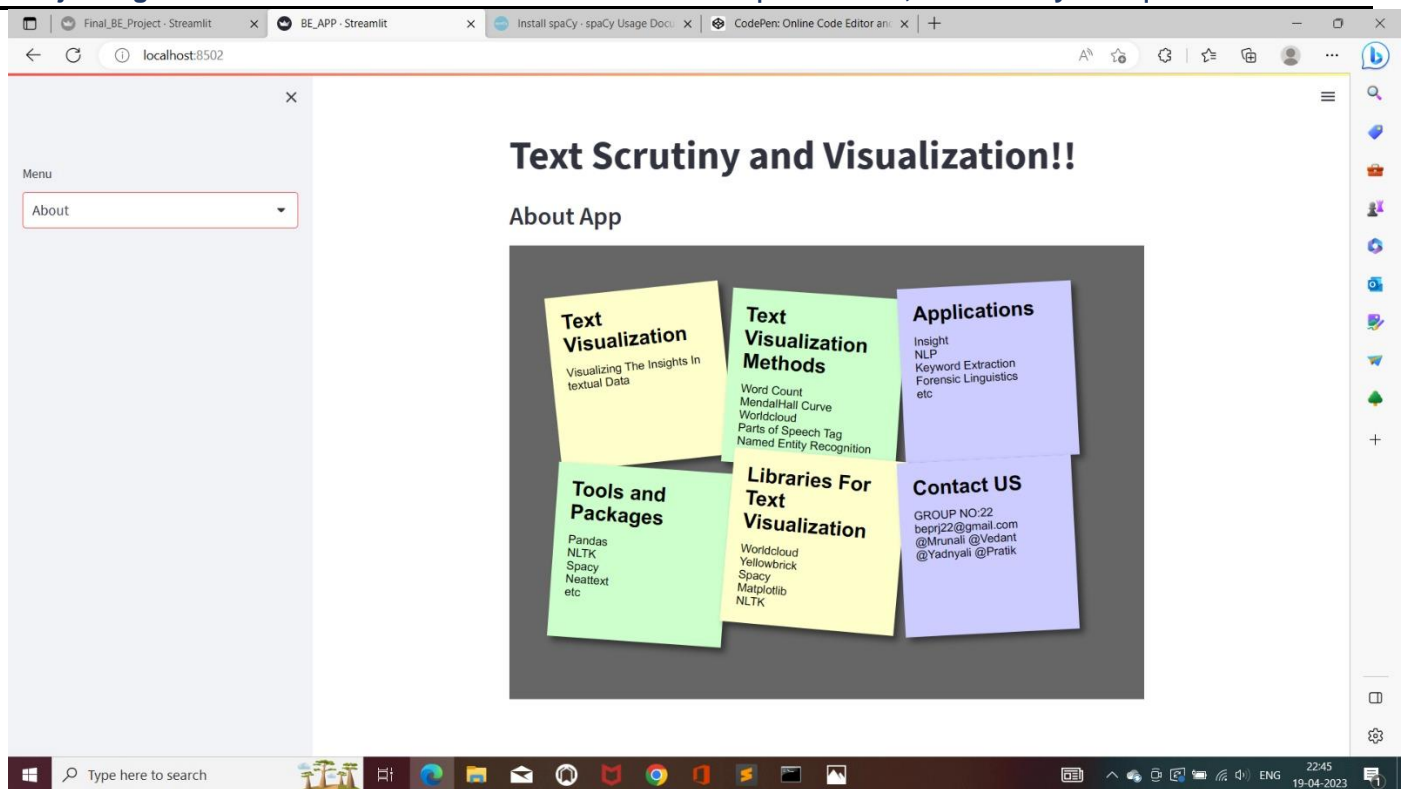


Figure 5: About Section

## IX. FUTURE WORK

This paper's main goal was to illustrate the importance of text analysis and visualization. As we have discovered how important a role the system plays in the contemporary world, we may further integrate this system on massive datasets with improved algorithms and use it in enterprises to make quick data-driven decisions. In the future, we will also permit the system to learn from regional tongues like Hindi and Sanskrit. This would enable the researchers to gain crucial insights from the material written in this language and aid in resolving issues facing the modern world.

## X. CONCLUSION

The study's findings support the notion that natural language processing (NLP) techniques are extremely advantageous and will be very useful for people who must read and evaluate regularly. Additionally, the user may quickly and concisely explain the results to administrators. The project fulfils its commitment to extract the important information, provide a visual representation, and present it in a way that provides the information with important insights.

## REFERENCES

- [1] Sarthak J Shetty , Vijay Ramesh, “ An open-source Python package for scientific text analysis” , DOI: /10.1002/ece3.8098.
- [2] Paritosh D. Katre, “ NLP Based Text Analytics and Visualization of Political Speeches” , ISSN: 2277-3878, Volume-8 Issue-3, 10.35940/ijrte.C6503.098319.
- [3] Akshaya Udgave, Prasanna Kulkarni “ Text Mining and Text Analytics of Research Articles” , (ISSN 1567-214x ), PJAEE, 17 (6).
- [4] John Risch, Shawn J bohn, Anne kao, Steve Potet “ Text Visualization”,DOI10.1007/978-1-4899-7502-7\_837-1.
- [5] Diksha Khurana, Aditya Koli, Kiran khatter & Sukhdev Singh “ Natural language processing: state of the art, current trends and challenges”, 10.1007/s11042-022-13428-4.
- [6] YunYun Yang, Lucy Akers, Thomas Klose, Cynthia Barcelon, Yang “Text mining and Visualization Tools [World Patent Information 30(2008) 280-293] BristolMyers Squibb, P.O. Box 4000, Princeton NJ 08543-4000 USA.