

## ARTICLE

# REGIONAL LANGUAGE DRIVEN PRESENTATION TOOL

Sharayu R Puranik\*, D. M. Thakore

Department of Computer Science, Bharti Vidyapeeth (Deemed to be University) College of Engineering, Pune,  
INDIA

### ABSTRACT

**Background:** India is a Big research hub for Natural Language Processing area. There are 22 Regional languages in India which can be worked upon for various aspects of processing. Research done in Natural language processing area will always be interdisciplinary at the border between Linguistic and Artificial Intelligence. **Methods:** This paper aims at providing details about a regional language driven presentation tool (RLDPT), which will take user input in the form of Sanskrit Nyasa (from Mathematical Grantha Lilavati) and would convert that mathematical expression into algorithm followed by its Graphical presentation. Tokenization, Noise removal, Entity extraction methods are called during processing. **Results:** Once user enters Nyasa for a specific mathematical method, the mathematical method is represented in Graphical format and user can run the process for sample numbers. **Conclusion:** The RLDPT tool is developed to facilitate the easy and visual learning of Sanskrit Nyasa. The paper has taken Sanskrit Grantha Lilavati as a base and would represent the mathematical expressions written in Sanskrit into English language algorithm. The APIs and algorithm would be used for other areas of Sanskrit literature. Also this can be extended for other Indian languages.

### INTRODUCTION

Having 22 Regional Languages, India is always treated as big research hub for Natural Language Processing area. Aim of Natural Language Processing and Artificial Intelligence area is to develop computer programs capable of human-like behavior related to 'understand given texts or produce meaningful texts' [1] in natural languages such as Sanskrit, Marathi, Hindi, English and many other regional languages. The most important applications of natural language processing include Retrieval of Information [1], Organization of the collected information, Machine Translation, Automatic Summarization, Sentiment Analysis, Text Classification and many more.

As in any science, activities of researchers are manly concentrated on its internal art and craft. Many problems arise during analysis and generation of Natural Language texts. Researchers focus on the solution of these problems of Semantic and Syntactic analysis, compilation of dictionaries, language text and grammar ambiguities.

In this project, we map the Nyasa written in Sanskrit (from Lilavati Grantha [2], into corresponding Algorithm steps and its Graphical representation. This would clearly show the mathematical steps mentioned in that specific method.

Being Emerging field in India, Natural Language Processing (NLP) [3] has a very good potential for research. Extensive research can be done in below NLP levels;

1. 'Conversion from Speech to text' technology,
2. Understanding of the Natural Language Text (with context) and
3. Effective and Efficient management/organization of the knowledge[3]

The main objective is to develop applications which are more relevant to those people speaking various regional languages.

Natural Language Processing (NLP) is a field of computer science, artificial intelligence and computational linguistics concerned with the interactions between computers and human (natural) languages [4]. The idea of using a natural language for computer programming is to make it easier for people to talk to computers in their native languages. For many, it is tedious and painful to learn Computer friendly languages like assembly, C, C++, Java, LISP etc. Use of native languages for Computer programming relieves such pain of learning Computer languages [4].

Multiple languages are spoken in India, each with its own flavor. Being mother of all languages, Sanskrit is the perfect language for computer programming. This language is grammatically perfect and has huge treasure of knowledge from all the fields [5].

Among all the Natural Languages, Sanskrit in its style is identified to be the best language which has minimum deviation. The creator of Sanskrit grammar, Panini, formulated 3,949 rules. Sanskrit is said to be a Mother of all languages. It deals with multiple limitations of Artificial Intelligence like NLP, Semantic Net, Vibhakti, Dual Case, Inflection based Syntax etc. Sanskrit language fulfils almost all of the prerequisites of a Natural Language Processor [5].

Information retrieval and information organization are the most important applications of natural language processing.

#### KEY WORDS

Natural Language Processing, Artificial Intelligence, Sanskrit, Lilavati

Received: 14 May 2018  
Accepted: 1 June 2018  
Published: 4 June 2018

#### \*Corresponding Author

Email:  
sharayu\_mirasdar@rediffmail.com  
Tel.: +91-8308609600

Some other applications are Natural Language Interfaces, Machine Translation and many more. In Natural Language Processing, below eight Technical areas can be considered for both theoretical study and application development:-

- Information Retrieval and Text Clustering
- Morphology, Syntax, Named Entity Recognition
- Semantics
- Opinion, Emotions, Textual Entailment
- Text and Speech Generation
- Machine Translation
- Educational Applications
- Applications

We can apply NLP techniques to retrieve treasure of knowledge, written by our ancestors, in Sanskrit.

Bhaskaracharya wrote Siddhantashiromani at the age of 36, Lilavati is the first part of it. The main Grantha Siddhantashiromani consists of four parts namely (Bhaskaracharya: 1144 – 1223 AD).

- 1) Lilavati (लीलावती)
- 2) Algebra (बीजगणित)
- 3) Planetary motions (ग्रहगणित)
- 4) Astronomy (गोलाध्याय).

Lilavati, the first 'prakarana' of Siddhantashiromani deals with 'Pati-Ganit' i.e. 'VyaktaGanit' or Arithmetic in today's Mathematical Term. It contains 278 verses. Being a Kavi also, Bhaskaracharya has written these verses in Poetic form (Shlokas) in Sanskrit language. There are certain verses which deal with Mensuration (measurement of various Geometrical Objects), Volume of Pyramids, Cylinders, heaps of grains etc., wood cutting, shadow and trigonometric relationship. Also on certain elements of Algebra such as finding an unknown quantity subject to certain constraints with the help of supposition method.

The Lilavati consists of 279 verses of rules and examples. The main contents are:

- Basic arithmetic operations including square roots and cube roots calculation for numbers, fractions, and the effect of text encryption.
- The rule of three, rule of five and so on
- Bartering, buying and selling
- Permutations and combinations
- Progressions and series
- Geometrical operations
- Solutions to indeterminate equations

In proposed system, we consider the Nyasa written in Lilavati Grantha. These Nyasa represent specific Mathematical formulae written for specific Mathematical Methods like Addition, Subtraction and so on.

Proposed system takes Sanskrit Nyasa as an input from front end GUI. Maps these Nyasa tokens into corresponding English words (Mathematical operations) and prepare an algorithm (in English) for the given method. The algorithm then also would be represented in the form of Flowchart. This would help user to clearly visualize the mathematical steps mentioned in that specific Sanskrit Nyasa.

## MATERIALS AND METHODS

The system consists of 6 modules in total.

### Modular design of a system

The first module is used to take input in form of Sanskrit Nyasa from user. User selects the type of mathematical expression for which Algorithm would be generated.

Types of mathematical formulas are

- गुणनेकरणसूत्रम् for Multiplication methods,
- भागाहरिकरणसूत्रम् for Division methods,
- वर्गकरणसूत्रम् to find Square of a number,
- वर्गमूलेकरणसूत्रम् for finding Square root of a number,
- घनेकरणसूत्रम् for finding Cube and
- घनमूलेकरणसूत्रम् to find Cube root of a number.

Second and third module deals with aspects of Natural Language Processing namely [6][7]

1. Removal of Noise from Input String (Removal of unwanted tokens)
2. Lexical Normalization (extract the exact word which may have multiple representations)

Entities are defined as the most important chunks of a sentence. Next module will extract entities from given Nyasa. Using look-up database table [Table 1], the module will find out mathematical operators from Nyasa.

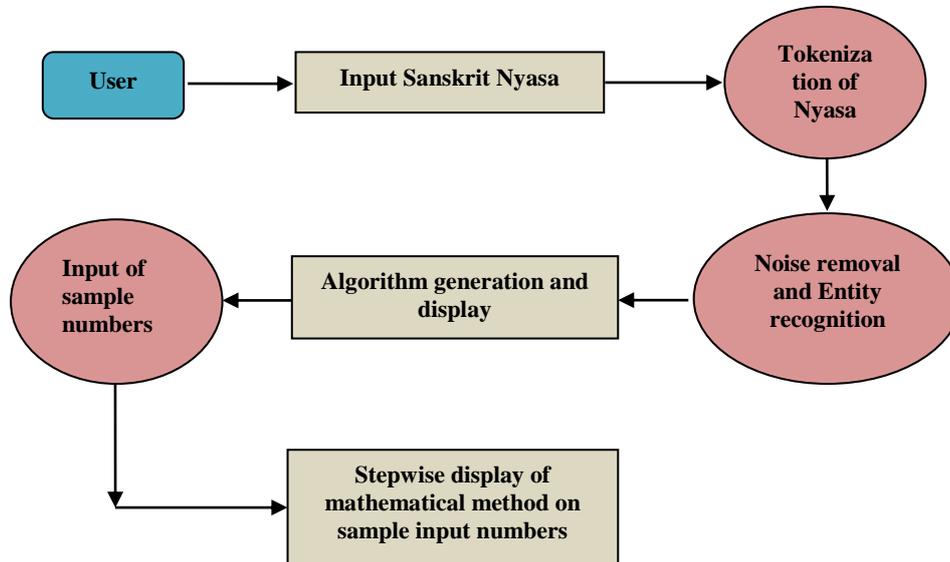
Module 5 and 6 deals with preparation of algorithm and execution of the method on sample input numbers.

**Table 1:** Lookup mapping table for Sanskrit-> English language words

Sanskrit Word	English Word	Operator
अंकं	Numbers	N
अन्त्यम्	Last	N
अन्वित	Addition	Y
उत्सारीतिन्	Remainder	N
ऊन	Subtraction	Y

Architectural diagram

Architectural diagram for the system [Fig. 1] shows various modules and their connection with each other.



**Fig. 1:** Architectural diagram of proposed system, Regional Language Driven Presentation Tool (RLDPT).

## Method

APIs are developed to process the Sanskrit (Regional Language) statements for Tokenization, Lexical Normalization and Entity Recognition [6][7]. These can be used for the other Regional language, provided with the database for that specific language.

For the generation of English language mathematical steps mentioned in Sanskrit Nyasa, Divide and Conquer algorithm is used. Once the mathematical operator is found from the Nyasa, the Nyasa is 'Divided' into left and right side of the operator. The operations to be performed on left and right side operands are formed stepwise. The final method of calculation is extracted and that operation is performed as 'Combine' process.

## RESULTS

The tool was run to generate algorithm for all 6 types of Mathematical methods. Sanskrit Nyasa was entered in the text box provided by the system. Algorithm got generated stepwise, in English language and displayed. The results after system run is shown in [Fig 2].

The algorithms then tested for sample input and results were studied for steps mentioned in Sanskrit Nyasa.

Form1

**॥ गणेशस्तुतिः ॥**

उपवासपूर्व-मङ्गल-परायम्—  
श्रीतिं भक्तजनस्य यो जनयते विघ्नं विनिघ्नन् स्मृत-  
स्तं इन्द्राकण्डवन्दितपदं नरस्य मनोज्ञानसम् ।  
पादं सद्गणितस्य वक्ष्ये चतुरश्रीतिप्रदां प्रसृष्टां  
संक्षिप्तश्रु-श्रीमला-ऽमलपदैर्लाजित्यतीतावतीम् ॥ १ ॥

Select type of करणसूत्रम्

गुणने करणसूत्रम्  
 आगाहारे करणसूत्रम्  
 वगे करणसूत्रम्  
 वर्गमूले करणसूत्रम्  
 घने करणसूत्रम्  
 घनमूले करणसूत्रम्

Enter the Sanskrit Nyasa (न्यास) in the TextBox below

गुण्य अन्त्यम् अङ्कम् गुणकेन इत्यात् उत्सारितेन एवम् उपान्तिस आदीन्

Show Graphical representation

Form2

Mathematical Operaion : **Multiplication**

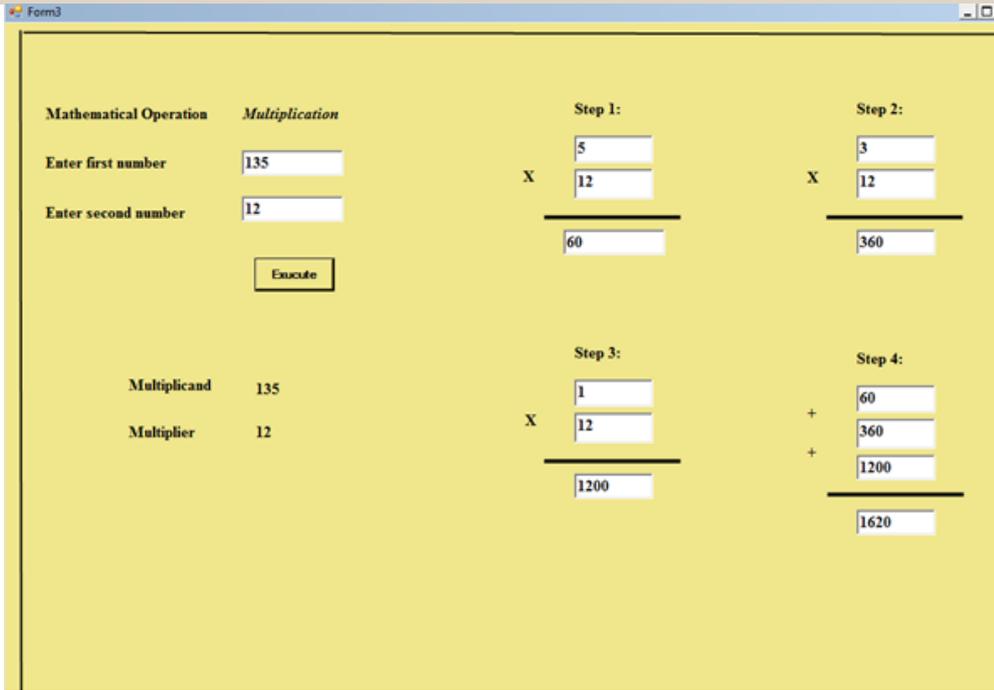
Description : **Method for Multiplication of two numbers**

Algorithm:

1. Take last digit (Unit place) of multiplicand
2. Multiply it with multiplier number
3. Place the answer at unit place (i.e. multiply answer by 1 and place)
4. Goto next digit (of 10th place)
5. Multiply that number with multiplier and place result at 10ths place (i.e. multiply answer by 10)
6. Repeat the process till all the digits of multiplicand are considered
7. Add all the numbers of Unit, Tenth, Hundreds etc. place
8. Print the final answer

If you want to execute above algorithm with Some example, click on the button 'Try Example'

Try Example



The screenshot shows a web-based interface for a mathematical operation. On the left, under 'Mathematical Operation Multiplication', there are input fields for 'Enter first number' (135) and 'Enter second number' (12), and an 'Execute' button. Below this, the multiplicand (135) and multiplier (12) are displayed. The main area shows four steps of the multiplication process:

- Step 1:** 5 x 12 = 60
- Step 2:** 3 x 12 = 360
- Step 3:** 1 x 12 = 1200
- Step 4:** The sum of the previous steps: 60 + 360 + 1200 = 1620

Fig. 2: 'RLDPT system run' Screen shots.

## DISCUSSION

The research done in this paper started with the aim to do work on Indian Regional Languages. Being mother of all languages, Sanskrit is the perfect language for computer programming. This language is grammatically perfect and has huge treasure of knowledge from all the fields.

Extant manuscripts in Sanskrit number over 30 million - one hundred times those in Greek and Latin combined - constituting the largest cultural heritage that any civilization has produced prior to the invention of the printing press. Sanskrit works include extensive epics, subtle and intricate philosophical, mathematical, and scientific treatises, and imaginative and rich literary, poetic, and dramatic texts.

Large digital platforms such as Google, Microsoft, YouTube, Facebook, C-DAC have stepped up their efforts to engage many of the Indian languages. The research in this paper is done considering Sanskrit languages. The Nyasa written in Sanskrit language, for mathematical domain, are processed on various aspects of Natural Language Processing and then presented as an algorithm and graphical representation [8] for that mathematical formula with sample input. As the research on Indian languages still in progress, separate APIs are built to carry out various activities on Sanskrit Nyasa. There is no existing system which can take Sanskrit as input language and translate the mathematical methods into Algorithms and visual presentation of it. The system is built from scratch to process Indian Languages (Sanskrit taken as base language) for NLP aspects.

## CONCLUSION

The objective behind development of a Regional Language Driven presentation tool can be stated as

- Help Sanskrit Language Learners with Graphical Representation tool, for Mathematical Shlokas.
- Extract vast Mathematical domain knowledge present in the great Indian Language 'Sanskrit'.
- Preserve culture, heritage and literature of the Indian languages.

The paper proposes a Regional language tool (Which is a Sanskrit Language) which would accept input Nyasa from Lilavati Grantha and would represent the mathematical methods mentioned in it graphically. This tool can be used by Learners of Sanskrit Language and can be extended for other Indian Languages in which the mathematical formulae are mentioned. The same can further be extended for other learning areas like Science, Economics, Politics and many other where ancient Indian literature is present in Regional Languages.

### CONFLICT OF INTEREST

None

### ACKNOWLEDGEMENTS

None

### FINANCIAL DISCLOSURE

None

**REFERENCES**

- [1] Mary Elaine Califf, Raymond J Mooney. [1997] Relational Learning of Pattern-Match Rules for Information Extraction
- [2] MD Pandit. [2013] English translation of Sanskrit Grantha Lilavati.
- [3] Ralph Weischedel, [1998] Chairperson BBN Systems and Technologies Corporation ,Jaime Carbonell Carnegie-Mellon University Barbara Grosz Harvard University ,Wendy Lehnert University of Massachusetts, Amherst Mitchell Marcus University, of Pennsylvania Raymond Perrault SRI International ,Robert Wilensky University of California, Berkeley, White Paper ; Natural Language Processing.
- [4] RadaMihalcea, Hugo Liu, Henry Lieberman; [2006] Natural Language Processing for Natural Language Programming
- [5] Pawan Goyal, Gérard Huet, Amba Kulkarni, Peter Scharf, Ralph Bunker; [2012] A Distributed Platform for Sanskrit Processing. DOI: 10.1.1.307.653
- [6] Cynthia A. Thompson, Raymond J.[1998] Mooney TR AI98-273, Artificial Intelligence Lab, University of Texas at Austin, Semantic Lexicon Acquisition for Learning Natural Language Interfaces.
- [7] Konrad Rieck, Christian Wressnegger. [2016] Journal of Machine Learning Research Homepage Harry: A Tool for Measuring String Similarity. 17(9):1-5
- [8] Jochen M. Kuster, Jana Koehler, Ksenia Ryndina; [2006] Improving Business Process Models with Reference Models in Business-Driven Development.