Exploring Ruby Based Bio-informatics R&D Framework Using NLP/BioNLP/SVM/QRNG/HPC/IoT/Mongo DB/BaseX DB Systems & Related Environments – A Novel Multidisciplinary Approach.

[Probing Advanced Computational Aspects of Cancer Research With Mathematics]

Nirmal Tej Kumar

Independent Consultant Informatics/AI/Photonics/Imaging/Nanotechnology/HPC R&D.
R&D Collaborator USA/UK/Israel/Germany/Japan/BRICS Group of Nations.
Current Member ante Inst,UTD,Dallas,TX,USA.
email id hmfg2014@gmail.com

Abstract :

An Insight into Probing [Ruby-NLP/BioNLP/qrng lib-mruby/rb-libsvm-Support Vector Machines/QRNG-Device/QRNG-Services Tool Kit/DNA/RNA Seq based Cancer Genetics Informatics Framework - Implementing Novel Algorithms] in Designing Future Bio-informatics R&D in the Context of [Ruby language/IoT/HPC/LLVM/LLVM-rb/] Heterogeneous Environments.

Keywords – NLP/BioNLP/SVM/QRNG/DNA Seq/LLVM/IoT/HPC/Ruby language

NLP – Natural Language Processing

Bio NLP – Biomedical Natural Language Processing

SVM/rb-libsvm -rb-libsvm -- Ruby language bindings for LIBSVM./Support Vector Machines.

DNA Sequencing - DNA sequencing is the **process of determining the nucleic acid sequence** – the order of nucleotides in DNA. It includes any method or technology that is used to determine the order of the four bases: adenine, guanine, cytosine, and thymine.

LLVM - The **LLVM** Project is a collection of modular and reusable compiler and toolchain technologies. Despite its name, **LLVM** has little to do with traditional virtual machines – Low Level Virtual Machines.

IoT- Internet of Things.

HPC- High Performance Computing.

QRNG Service - Quantum Random Number Generator Service

Ruby Language - Ruby is an interpreted, high-level, general-purpose programming language. It was designed and developed in the mid-1990s by Yukihiro "Matz" Matsumoto in Japan.

[I] Inspiration & Introduction :

" Interest in natural language processing (NLP) began in earnest in 1950 when Alan Turing published his paper entitled "Computing Machinery and Intelligence," from which the so-called Turing Test emerged. Turing basically asserted that a computer could be considered intelligent if it could carry on a conversation with a human being without the human realizing they were talking to a machine.

The goal of natural language processing is to allow that kind of interaction so that non-programmers can obtain useful information from computing systems ".

Alissa Lorentz Writes - "resulting in a data mining gold rush that will soon have companies and organizations accruing Yottabytes (10^24) of data." ["With Big Data, Context is a Big Issue," *Wired Innovation Insights*, 23 April 2013] NLP is very important".

[Source : https://www.wired.com/insights/2014/02/growing-importance-natural-language-processing/]

Some info from the Literature sources that could be useful "why" we wish to use QRNG Device/Software/Libraries in our R&D in Processing Information Using NLP Techniques :

[a] Brontobyte (10^27) of data.

[b] Yottabytes (10^24) of data.

[c] Exabyte (10^18) of data .

[d] The petabyte is a multiple of the unit byte for digital information. The prefix peta indicates the fifth power of 1000 and means 10^{15} in the International System of Units (SI), and therefore 1 petabyte is one quadrillion (short scale) bytes, or 1 thousand billion (long scale) bytes. The unit symbol for the petabyte is PB.

[e] The terabyte is a multiple of the unit byte for digital information. The prefix tera represents the fourth power of 1000, and means 10^{12} in the International System of Units (SI), and therefore one terabyte is one trillion (short scale) bytes. The unit symbol for the terabyte is TB.

Nucleic Acids Data Sequencing using Higher Order Logic-A Suggestion of Basic Computational Framework Towards Bio-Sensors and Gene-Chips Design, Implementation and Verification* - D.N.T.Kumar et al 2012.
[Source - http://www.scienpress.com/Upload/JAMB/Vol%202_2_6.pdf]

https://github.com/zjms/allennlp - An open-source NLP research library, built on PyTorch. http://www.allennlp.org/
[Source - https://allennlp.org/ - AllenNLP is a free, open-source project from AI2.]
AI2 - The Allen Institute for Artificial Intelligence.
https://nlp.stanford.edu - NLP Group - Stanford Univ,California,USA.

Here,we are not focusing on Python based NLP stuff. We are focusing only on Ruby Language for NLP & Informatics.

[f] https://www.lifewire.com/terabytes-gigabytes-amp-petabytes-how-big-are-they-4125169

[g] https://chewychunks.wordpress.com/2013/03/23/future-of-big-data-structure/

[h] *https://www.ibm.com/developerworks/library/os-createcompilerllvm1*

[i] *llvmweekly.org* && *https://**repl.it**/languages/**Ruby*** && http://mind.sourceforge.net/ruby.html

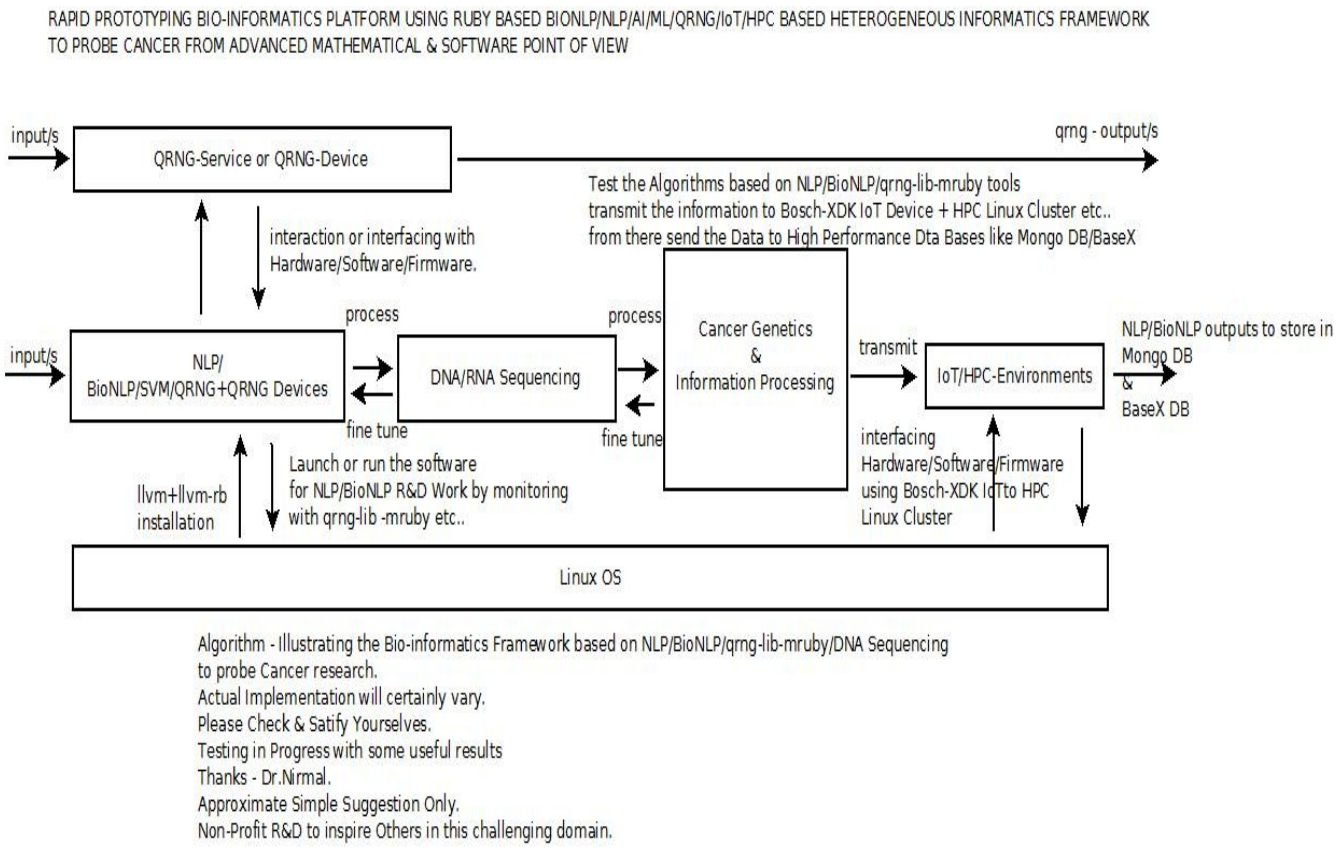
[j] *https://**github.com**/kokubun/llrb* && ***mruby**.org* && *www.**ruby**inside.com/**artificial-intelligence-ai-programming-in-ruby**-1157.html*

[k] *https://github.com/mruby/mruby* && https://qrng.physik.hu-berlin.de *****

[l] http://scirubyj.com/ - Tools for Scientific Computing in Ruby && https://ruby.ai/

[m] https://github.com/febeling/rb-libsvm – Ruby based SVM-Support Vector Machine.

[II] BioNLP/NLP/qrng-lib mruby/rb-libsvm/DNA Seq based R&D Bio-Informatics Framework :



Algorithm - Illustrating the Bio-informatics Framework based on NLP/BioNLP/qrng-lib-mruby/DNA Sequencing to probe Cancer research.
Actual Implementation will certainly vary.
Please Check & Satify Yourselves.
Testing in Progress with some useful results
Thanks - Dr.Nirmal.
Approximate Simple Suggestion Only.
Non-Profit R&D to inspire Others in this challenging domain.

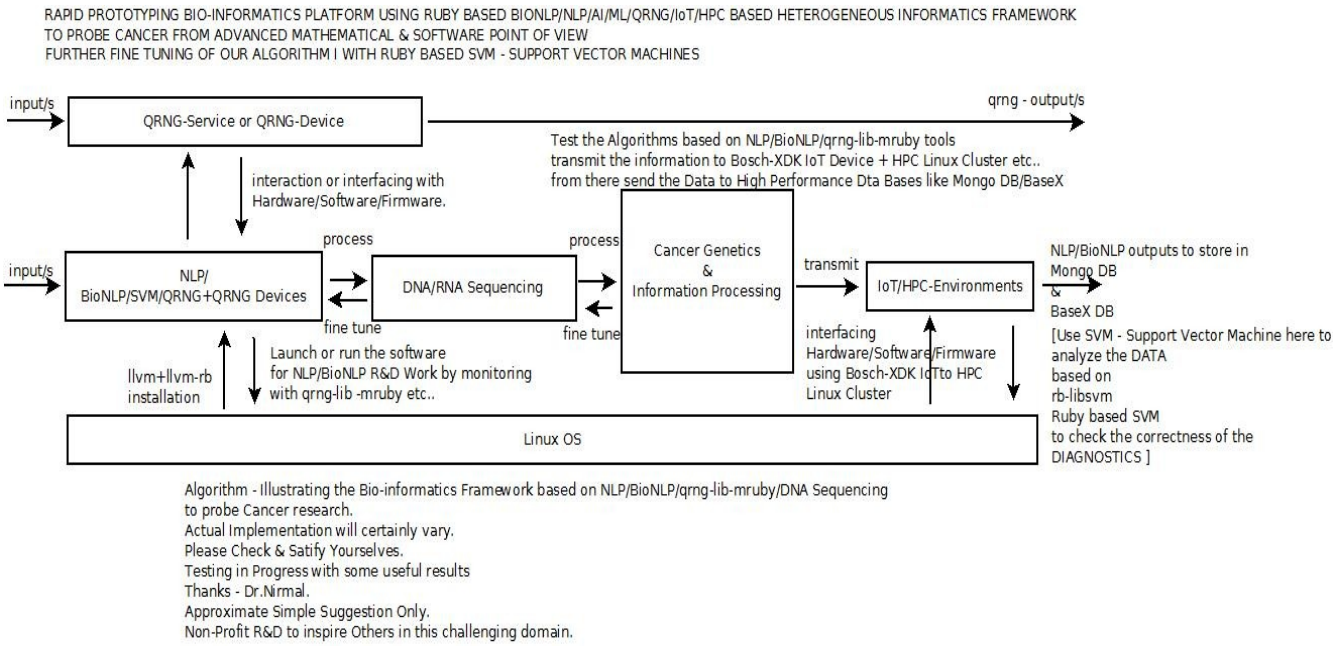
[Figure I- Algorithm I – Our R&D Bio-informatics Framework Using Ruby based Software Tools]

http://www.vixra.org/author/nirmal_tej_kumar (((via))) Vixra.org - Some useful information.

"Cancer is a genetic disease—that is, cancer is caused by certain changes to genes that control the way our cells function, especially how they grow and divide. Genes carry the instructions to make proteins, which do much of the work in our cells". [Source - *https://**www.cancer.gov**/about-**cancer**/causes-prevention/**genetics***]

http://michorlab.dfci.harvard.edu/publications/Naturereviewscancer2015.pdf – **Mathematics of Cancer – Interesting.**

http://rubynlp.org/ - Awesome NLP with Ruby - Useful resources for text processing in Ruby .



[Figure II – Algorithm II – Fine Tuning of Algorithm I With Ruby based SVM – Support Vector Machines]

[III] Related R&D Information on Mathematics+Software Used :

[a] https://www.academia.edu/1171399/Information_Processing_in_Random_Automata_Networks

[b] https://www.idquantique.com/random-number-generation/overview/

[c] https://nlp.stanford.edu/nlaspAbs/jelinek.shtml

[d] https://github.com/cremno/mruby-libqrng

[e] http://rubynlp.org/

[f] https://www.nature.com/articles/s41586-018-0559-3

[g] http://www.vixra.org/author/nirmal_tej_kumar

[h] *https://www.**ruby**-lang.org* - **Ruby Programming Language**

[i] *https://**github.com/ruby-llvm/ruby-llvm***

[j] *https://**llvm**.org* – LLVM

[k] *https://**stackoverflow.com**/questions/2354725/**what-exactly-is-llvm***

[l] *https://**github.com/llvm*** - LLVM

[m] *https://riptutorial.com/**llvm***

[n] *https://**clang.llvm**.org*

[o] *https://apt.**llvm**.org*

[p] *https://www.analyticsvidhya.com/blog/2019/03/pretrained-models-get-started-**nlp***

[IV] Acknowledgment/s :

Special Thanks to all WHO made this happen in my LIFE. Non-Profit R&D.

[V] References :

[a] *https://**www.ncbi.nlm.nih.gov**/pmc/articles/PMC4615216*

[b] *https://bio.fandom.com/wiki/**NLP***

[c] *https://towardsdatascience.com/ai-**in-bioinformatics**-a1acdc3cdd89?source=rss----7f60cf..*

[d] http://isoft.postech.ac.kr/Course/NLP_for_Bioinformatics/index.html

[e] http://vixra.org/pdf/1905.0186v1.pdf – Important R&D Highlighting our Research.

[THE END]