INPE – National Institute for Space Research
São José dos Campos – SP – Brazil – May 16-20, 2016

A WEB FRAMEWORK FOR ADVANCED AND INTENSIVE NONLINEAR TIME SERIES ANALYSIS

Bruno B. F. Leonor[1], *Walter A. dos Santos*[2], *Asiel B. Jr*[3], *Reinaldo R. Rosa*[4]

[1]National Institute for Space Research - INPE-LAC-CAP, São José dos Campos, Brazil, brunobfl@yahoo.com.br
[2]National Institute for Space Research - INPE-DEA-ETE, São José dos Campos, Brazil, walter.abrahao@inpe.br
[3]National Institute for Space Research - INPE-CPTEC, São José dos Campos, Brazil, asiel.bomfin@cptec.inpe.br
[4]National Institute for Space Research - INPE-LAC, São José dos Campos, Brazil, rrrosa.inpe@gmail.com

Abstract: Using the e-Science paradigm, computers play a key role in supporting scientific areas such as those using nonlinear time series analysis to understand seemingly unpredictable behavior. This work presents a web framework supporting researchers in this class of analysis in a cloud environment where they can access, parameterize, initialize and monitor their applications transparently. Moreover, new applications can be added and managed by researchers on the portal easily. As a case study, techniques such as detrended fluctuation analysis (DFA) and gradient pattern analysis (GPA) will be added and used for intensive time series analysis in the Brazilian Space Weather Program.

keywords: Time Series Analysis, Nonlinear Dynamics and Complex Systems, Chaos and Global Nonlinear Dynamics, Web Framework, High-Performance Computing.

1. INTRODUCTION

The web has been causing a growing impact on scientific research, technological development and society as a whole with its growth over the past few years and has brought a different style of computing [1].

Tony Hey, vice president of Microsoft Research during the Latin American e-Science Workshop in 2013, said: ”Some technologies used in computer science will help solve scientific problems. In contrast, the use of these tools to solve scientific problems also make possible the very development of computer science” [2].

This relationship with computing various scientific areas are being benefited with the use of computing environments increasingly modern. This has contributed to the creation of a powerful compound ecosystem of humans and machines that together are able to solve problems that could not be solved alone.

This paper presents an e-Science environment named Sentinel to support scientists in their research in a cloud environment where they can add, parameterize, initialize and monitor their applications transparently using any enabled device to access the web.

As a case study, techniques such as detrended fluctuation analysis (DFA) and gradient pattern analysis (GPA) will be added and used for intensive time series analysis in the Brazilian Space Weather Program (EMBRACE).

2. E-SCIENCE

The term e-Science is commonly used to describe the development of software services infrastructure [3]. This enables access to remote facilities, distributed computing resources, information storage in dedicated databases, dissemination and sharing of data, results and knowledge.

Research in e-Science requires joint efforts and multidisciplinary, where computer scientists assist researchers from other fields of knowledge to develop research more quickly and efficiently and that this partnership, the computer is also made of innovative and transformative way [4].

The research in e-Science can be described by the following characteristics [4]: (1) involves collaboration of researchers computing with researchers from other fields of knowledge, and (2) requires the creation of sophisticated computational methods to handle large volumes of data (Big Data) and/or to perform simulations and programs that require complex computational systems.

3. SENTINEL

The environment proposed in this paper is named "Sentinel" and it is an e-Science environment that aims to provide high performance computing resources to scientists to develop their research. Scientists affiliated to the portal can register their applications and transparently manage them, more specifically, they can access, parameterize, initialize, monitor, analyze and share the results obtained through any device that allows access to the web.

There are two ways to access the portal, as a researcher or guest. The main features for a researcher are shown in Figure 1.

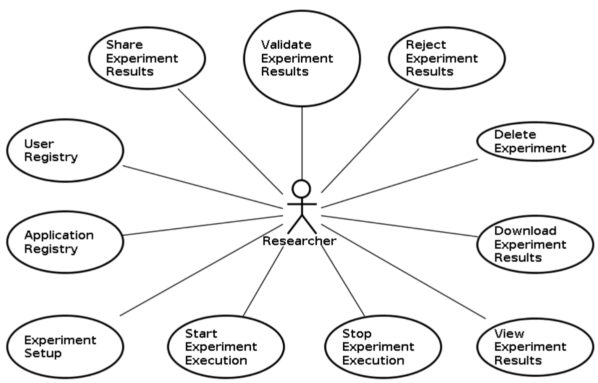

Figure 1 – Researcher Functions.

Most of these features are available for each of the experiments via the Dashboard interface in the Actions column, as shown in Figure 9. Each researcher will have you own desktop with your set of tools that will be built from the application registry interface (Figure 3).

A guest can only view and download a shared results as shown in Figure 2.

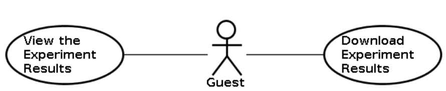

Figure 2 – Guest Functions.

The Sentinel portal allows researchers to concentrate all necessary applications to develop their research in a single environment without the need to meet the computing infrastructure used.

The development is being done using the Java language because it is the programming language most commonly used by developers around the world according to the rankings: Tiobe [5], RedMonk [6] e Spectrum [7].

3.1. Prototype

The prototype of the Application Registry interface is shown in Figure 3. In this interface the researcher will add all the information necessary for the implementation of application, ranging from the programming language used in the application setting up parameters needed for its execution. This informations are grouped into the tabs "Fields", "Files", "Libs" and "Compile".

Figure 3 – Application Registry Interface.

Each registered application may have their visibility set by the researcher to be either public or private.

The tab Fields is presented in Figure 4 which will define the input variables for the application parameterization. For each field, it will be assigned a type, a default value and description if necessary.

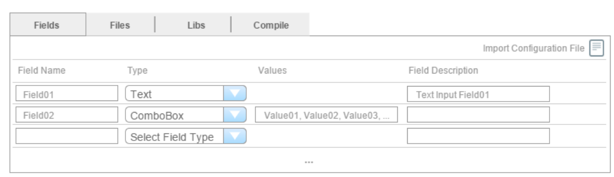

Figure 4 – Application Registry Interface - Tab Fields.

In the tab Files, see Figure 5, is intended to upload the application source code.

Figure 5 – Application Registry Interface - Tab Files.

If necessary, the use of some special library may be added to the environment by tab Libs as shown in Figure 6.

Figure 6 – Application Registry Interface - Tab Libs.

The last tab, "Compile", is used to define the application's build parameters within the Sentinel portal environment where the researcher must enter a script to compile and run the application.

The prototype of the Experiment Setup interface is shown in Figure 7. This interface will be created automatically by the framework based on the information entered in the application registry interface - Fields tab (Figure 4). The user needs to set the values that will serve as input for the execution of the application. Settings created to perform this step are generically called an experiment.

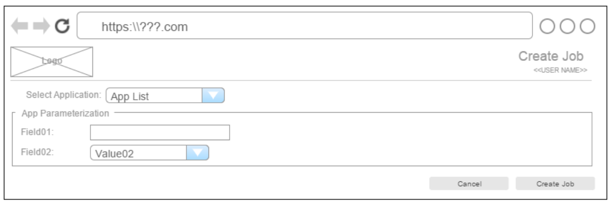

Figure 7 – Experiment Setup Interface.

The Dashboard interface, prototype shown in Figure 8, is where the user can track all your experiments.

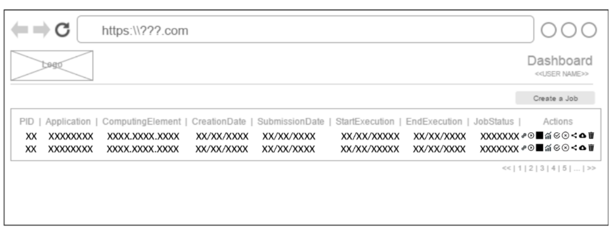

Figure 8 – Dashboard Interface.

Through the Dashboard interface, researchers will have access a set of informations about each experiment, which are:

PID - experiment identifier.

Application - chosen application.

ComputingElement - which node is running.

CreationDate - creation date of the experiment.

SubmissionDate - request date for execution.

StartExectution - initial date of execution.

EndExecution - end of the execution date.

JobStatus - current situation of the experiment.

Actions - actions to be taken by the researcher.

The features available in the "Actions" column into Dashboard interface are shown in the Figure 9, where the researcher may: Experiment Setup, Start Execution, Stop Execution, View Results, Validate Results, Reject Results, Share Results, Download Results and Delete Experiment.

Figure 9 – Experiment Actions.

To view the results, Sentinel portal will provide to researchers a set of tools for visualization that will provide charts, 2D and 3D plot, statistics, histogram, among others. After the experiment is visualized and evaluated the results may be accepted or rejected. Rejected results will be removed from the platform, accepted results will be available to the scientists for a time not set yet. These actions can be taken for each of the experiments.

4. CASE STUDY ON DFA

As a case study, techniques such as DFA (Detrended Fluctuation Analysis) [8] and GPA (Gradient Pattern Analysis) [9] will be incorporated into the Sentinel portal and used for analysis of time series obtained the Brazilian Space Weather Program, named "Embrace" [10].

This Space Weather program is mainly concerned with the area of knowledge related to solar phenomena and physical occurrences in the space environment [10], which are manifested in recurrent way and affect the stars and artifacts in space.

Time series are usually used as source of information for grasping phenomena understanding and scientific research in various sectors of society. Time series analysis assumes that set of the observations is being generated by a stochastic process and as such it has a probabilistic framework that can be characterized and described [11].

The time variability of many natural and social phenomena is not well described by standard methods of data analysis. However, nonlinear time series analysis comprises a wide variety of measures that allows one to extract different features of a dynamical system underlying some measured signal [12]. The results can be applied to real data from physics, biology, medicine, and engineering.

For analysis of the time series obtained by Embrace program it will be used two techniques, DFA and GPA.

DFA was basically designed to reveal long range correlation in non stationary processes. However DFA can also be regarded as a suitable method to investigate both long-range and short range correlation in non-stationary and stationary systems [8].

The Gradient Pattern Analysis is an innovative technique, which characterizes extended patterns based on large and small amplitude fluctuations of the spatial, temporal, and spatio-temporal structures represented as a static or dynamical gradient lattice [9]. The local fluctuation between each pair of amplitudes of the global pattern is characterized by its gradient vector at corresponding mesh-points in the two-dimensional space. In this representation, the relative values between adjacent amplitudes ("local fluctuations") are relevant, rather than the respective absolute values.

New techniques and algorithms may be incorporated into the Sentinel portal by scientists to complement their studies as well as new visualization tools.

5. CONCLUSION

This work presented a web framework named Sentinel dedicated and suitable for managing e-Science applications. It is expected that this work will contribute to a more environmentally friendly and safe for e-Science users, where they can gather all the applications used in the development of their research.

The Embrace program was chosen as a case study for the knowledge and prediction of different phenomena that directly affect human activities, such as solar activity, magnetic storms and ionospheric storms, they have great importance for the well being of society.

The framework has been applied for time series analysis using two techniques, DFA and GPA. Future work will incorporate more challenging experiments and algorithms.

References

[1] N. Shadbolt, W. Hall, J. A. Hendler, W. H. Dutton,"Web science: a new frontier", The Royal Society, vol. 371, n. 1987, 2013.

[2] E. Alisson, "eScience revoluciona a forma como se faz ciência.", available at: http://agencia.fapesp.br/17279, accessed January 2016.

[3] D. Palazzi, L. Silva, L. F. Mendes, W. Gaspar, E. Matos, F. Campos, R. Braga, "Uso de ontologias em projetos de e-science", BDBComp, 2009.

[4] Fapesp, "Programa FAPESP de Pesquisa em eScience", available at: http://www.fapesp.br/publicacoes/2015/folder_escience.pdf, accessed December 2015.

[5] Tiobe, "TIOBE Index for February 2016", available at: http://www.tiobe.com/tiobe_index, accessed February 2016.

[6] RedMonk, "The RedMonk Programming Language Rankings: January 2016", available at: https://redmonk.com/sogrady/category/programming-languages/, accessed February 2016.

[7] Spectrum, "The 2015 Top Ten Programming Languages", available at: http://spectrum.ieee.org/computing/software/the-2015-top-ten-programming-languages, accessed February 2016.

[8] V. V. Morariu, L. Buimaga-Iarinca, C. Vamos, S. Soltuz, "Detrended Fluctuation Analysis of Autoregressive Processes", Cornell University Library, 2007.

[9] R. R. Rosa, M. Karlický, T. B. Veronese, N. L. Vijaykumar, H. S. Sawant, A. I. Borgazzi, M. S. Dantas, E. B. M. Barbosa, R. A. Sych, O. Mendes, "Gradient pattern analysis of short solar radio bursts", Advances in Space Research, vol. 42, 2008.

[10] INPE, "O Programa Embrace", available at: http://www2.inpe.br/climaespacial/portal/o-programa-embrace/, accessed January 2016.

[11] R. M. Kirchner, R. C. Souza, F. A. Ziegelmann, "Identificação de estruturas não-lineares de séries temporais através de regressão linear local e modelos aditivos", Pesquisa Operacional, available at: http://dx.doi.org/10.1590/S0101-74382008000100003 vol. 28, 2008.

[12] H. Kantz, T. Schreiber, "Nonlinear Time Series Analysis", Cambridge University Press, 2004.