Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 4 (2011) 678–687

Procedia Computer Science

International Conference on Computational Science, ICCS 2011

A natural language programming solution for executable papers

Sandor M Veres[a]* and J. Patrik Adolfsson

Faculty of Engineering and the Environment
University of Southampton, Highfield, Southamton, SO17 1BJ, UK

Abstract

The paper describes a system for executable papers for publishers enabling them to reuse content and to generate further advances of science and engineering. The executable algorithmic descriptions within a paper are presented in natural language sentences and basic code, thereby making long term compatibility absolute. Authors are required to use publicly numerical libraries on the Internet or references to publications with executable papers. As used by authors the system automatically creates a web of algorithmic knowledge on the Internet. Novelty of new algorithms in publications can be evaluated by automated tools available to authors, reviewers and readers of scientific papers published.
Keywords: interactive publications, data driven computing, natural language programming

1. Objectives of the system proposed

The system described here is derived from machine readable and executable papers [1-5] in HTML/LaTex formats, for which a Reader Agent [5] is currently available. The system proposed has the following features:

1.1 It provides tools for creating natural language descriptions of executable algorithms (NLDEA) that can be included in PDF and HTML publications and which can be unambiguously interpreted/executed by a web-browser plug-in. Long term compatibility is absolute by use of natural language and self contained interpretation within the paper.

1.2 The NLDEA are formal to such a degree that they enable reviewers to use automated tools to compare past results published in this format. Past published and currently used ontologies and procedural steps can be compared for similarity. A similarity score is computed with 0 being totally dissimilar and 1 meaning essentially identical algorithms.

1.3 No grammar needs to be learnt by authors or readers of NLDEA. Sentence meanings are expressed by other sentences and sentence interpretation is by template fitting. High level computer language is only used for sentence meanings defined by algorithms in publicly available numerical libraries on the Internet.

1.4 The executable paper proposed is to be part of a wider system for machine intelligence where machines directly

* Corresponding author. Tel.: +44-023-80597754; fax: +44-2380-433122.
E-mail address: s.m.veres@soton.ac.uk.

Selection and/or peer-review under responsibility of Prof. Mitsuhisa Sato and Prof. Satoshi Matsuoka
doi:10.1016/j.procs.2011.04.071

read publications [6] to develop their intelligence. The proposed executable solution is however uch more than a solution to "executable" papers as asked by Elsevier. It leads to publications for machines where all the text of the paper is interpretable by intelligent agents. The system also allows full definition of intelligent agents in English sentences. Natural language programming (NLP) and machine readable sEnglish™ documents and editing tools were first publicly exhibited at the July 2008 IFAC World Congress in Seoul by SysBrain Ltd., UK.

This paper is limited to describing an executable paper system where the papers are either in PDF or in HTML formats and can contain one or several NLDEA sections. In this paper we identify an NLDEA section with an NLP section. Apart from NLDEA sections the rest of the paper is presented in the format of traditional human readable papers as specified by the respective publisher. The system described does not interfere with publisher's current paper presentational requirements. Executable sections can for instance be required by the publisher to be presented either in special fonts, boldface or italics.

2. Outline of the system

The system proposed and depicted within Fig. 1 is based on the inclusion of NLDEA in the form of natural language programming (NLP) text within a published paper, which may be interacted with to gain experience with published results. sEnglish™ is one type of NLP that uses English (www.system-english.com). A web-browser is used to view an electronic version of a published paper. Through the use of a browser plug-in, a reader can rerun experiments of the author or modify editable natural language descriptions of a numerical demonstration that is subsequently reinterpreted and demonstrated. The reader can also modify some natural language sentences to alter display modes of the results.

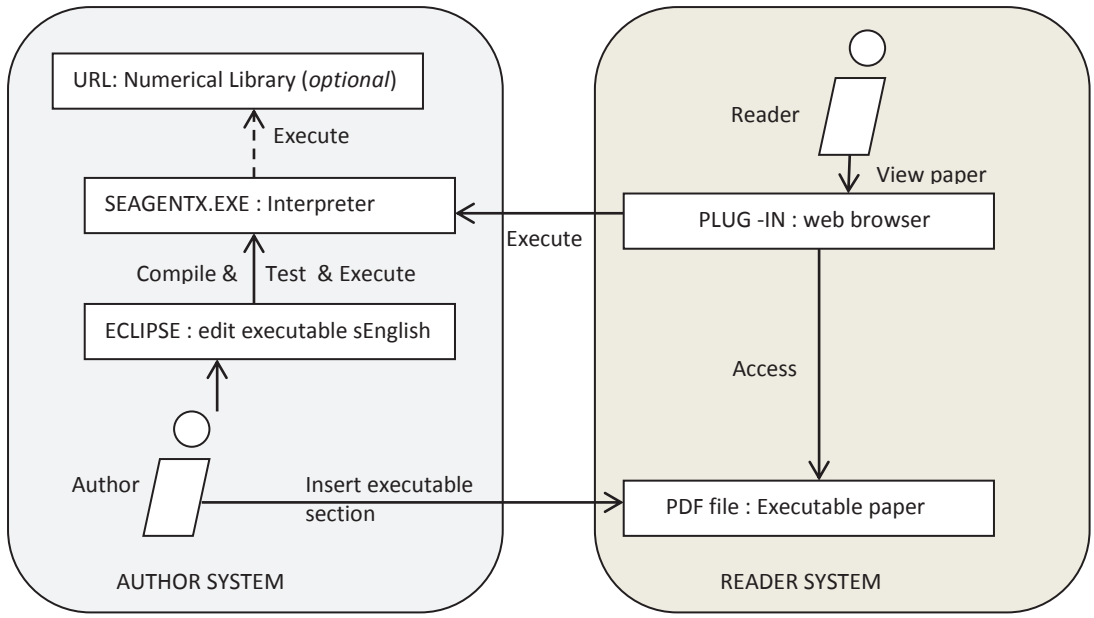

Fig. 1. Functionality of the executable paper system. Difference in [5] is the use of standalone programs not browser plug-ins

2.1 The appearance of executable sections in ordinary pdf publications

Fig. 3 illustrates an example of an NLP text in a paper. Our parser can unambiguously compiles these NLP sentences into executable MATLAB™ code. In a similar fashion, we have parser for Java based sentence meanings. There is the potential to expand NLP for the use of most popular computer languages: C++, Python and ADA, etc. Currently our Reader Agent [5] can use the full power of MATLAB's graphical data displays: the web-browser