Institut
"Jožef Stefan"
Ljubljana, Slovenija

TECHNOLOGY OFFER

ONLINE WORKFLOWS FOR DISTRIBUTED BIG DATA MINING

Fields of use

Databases and on-line information services, Big data management, Databases, Database Management, Data Mining, Knowledge Management, Process Management, User Interfaces, Usability

Current state of technology

Field tested/evaluated

Type of cooperation

Technical cooperation agreement, Research cooperation agreement

Intellectual property

Copyright, Secret Know-how

Developed by

Jožef Stefan Institute

Contact

Jožef Stefan Institute,
Jamova cesta 39,
1000 Ljubljana,
Slovenia
Phone: + 386 1 47 73 224
E-mail: tehnologije@ijs.si
Web site: http://tehnologije.ijs.si/

Summary

A Slovenian public research organization has developed a cloud-based platform that supports the composition and execution of data and text processing workflows. It fulfils the needs of many companies facing the problem of collecting huge amounts of data but lacking intuitive user-friendly data mining tools. The platform is provided as a hosted service, with the ability for users to install it on a private cloud.

Partners are sought for technical/research cooperation agreements.

Description of the invention

A Slovenian public research organization has developed a crowd-sourced workflows platform on the cloud. It is an open-source data processing workflow platform with a user-friendly graphical interface that can run in any browser and does not require installation on client computers. The platform offers software components supporting data analytics and pattern discovery and workflow components allowing graphical user interaction during runtime and visualization of results by implementing views in any format and rendered in a web browser. Data processing in the platform is managed by connecting processing components into a workflow executable on the cloud, and the graphical user interface for constructing workflows follows a visual programming paradigm that simplifies the representation of complex procedures into spatial arrangements of building blocks.

Similar and related platforms exist, but no workflow construction tool would match the ease of use and workflow sharing abilities of the platform. Currently, probably the most advanced workflow management system (used primarily for bioinformatics research) is conceived as a suite of tools used to design and execute scientific workflows. A multilingual Internet service platform for supporting Intercultural collaboration is based on a service-oriented architecture and supports a web-oriented version of the pipeline architecture typically employed by natural language processing tools. Yet another platform, a more recent development, enables workflows to have interactive components, where the execution of the workflow pauses to receive input from the user.

The basic unit of the platform proposed is a processing component, which is graphically represented as a widget. Considering its inputs and parameters every such component performs a task and stores the results as outputs. Different processing components are linked via connections through which data is transferred from a widget's output to another's input. Alternative widget inputs are parameters, which the user enters into a widget's text fields. The platform is easy-to-use, requires no expertise to construct new workflows while allowing complex data mining analyses to be performed on the input data set.

KONZORCIJ ZA PRENOS TEHNOLOGIJ IZ JRO V GOSPODARSTVO

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND

REPUBLIC OF SLOVENIA
MINISTRY OF EDUCATION, SCIENCE AND SPORT

The investment is co-financed by the Republic of Slovenia and the European Union under the European Regional Development Fund.

The researchers come from a Slovenian public research organization. Their technologies have been successfully applied to many practical problems, including earthquake prediction, selection of applicants for loans of the National Housing Fund, analysis of UK traffic accidents, medical diagnosis, analysis of the Slovenian public healthcare system, scientific digital editions of Slovene literature.

The partners are sought among data mining practitioners and developers for:
• Technical cooperation agreements in the implementation of existing workflows.

Research cooperation: development of new workflows.

Main Advantages

The graphical user interface (GUI) implements an easy-to-use formalism of arranging widgets on a canvas to form a graphical representation of complex procedures. Construction of new workflows requires no expertise apart from knowing (usually from widget documentation) the inputs and outputs of widgets to ensure their compatibility. Once constructed, the workflows can be shared (publicly or privately), reused, and extended.

Partner Sought

The partners are sought among data mining practitioners and developers.

• Technical cooperation agreements: implementation of existing workflows.

• Research cooperation agreements: development of new workflows in data mining and text processing applications.

KONZORCIJ ZA PRENOS TEHNOLOGIJ IZ JRO V GOSPODARSTVO

EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND

REPUBLIC OF SLOVENIA
MINISTRY OF EDUCATION, SCIENCE AND SPORT

The investment is co-financed by the Republic of Slovenia and the European Union under the European Regional Development Fund.