MANUAL DE BIBLIOTECA VIRTUAL SPRINGER

Ingresar a www.uagraria.edu.ec

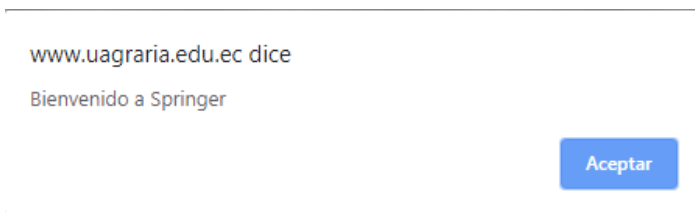Ir a la sección de SERVICIOS ONLINE – BIBLIOTECA VIRTUAL
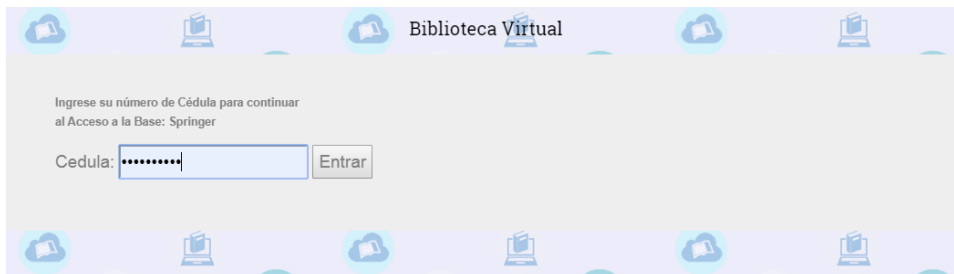

Seleccionar la biblioteca Springer

• Ingresar NÚMERO DE CEDULA (Docente o estudiante)

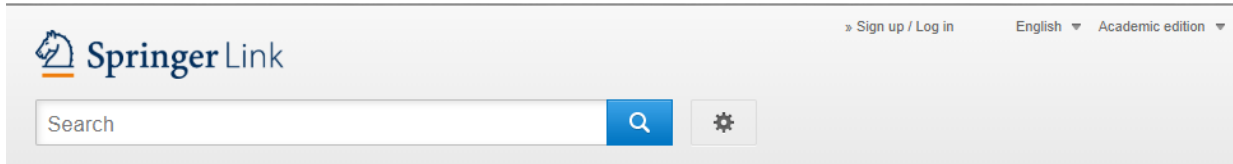o Sus accesos serán registrados para el informe

SPRINGER

Permite realizar búsquedas de: Articles, Chapters, Papers, Reference Works Entries, en varios idiomas pero principalmente en Ingles pero también incluye en Español.

- Para obtener mayores funciones es necesario registrarse en esta biblioteca y luego loguearse.
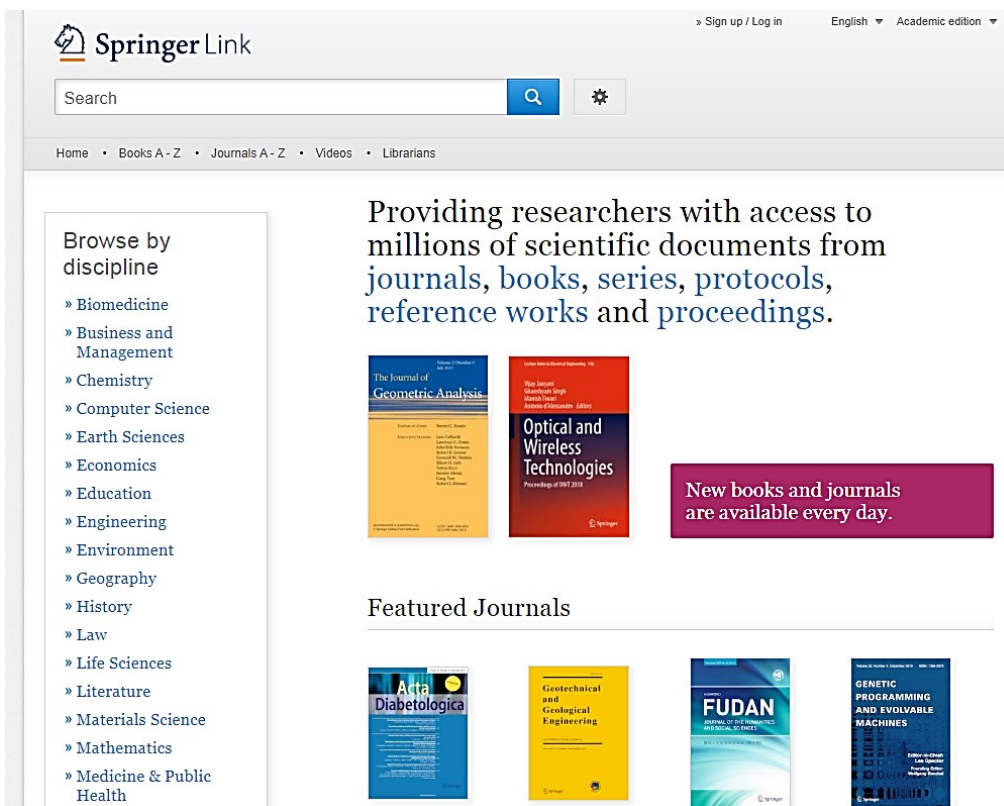
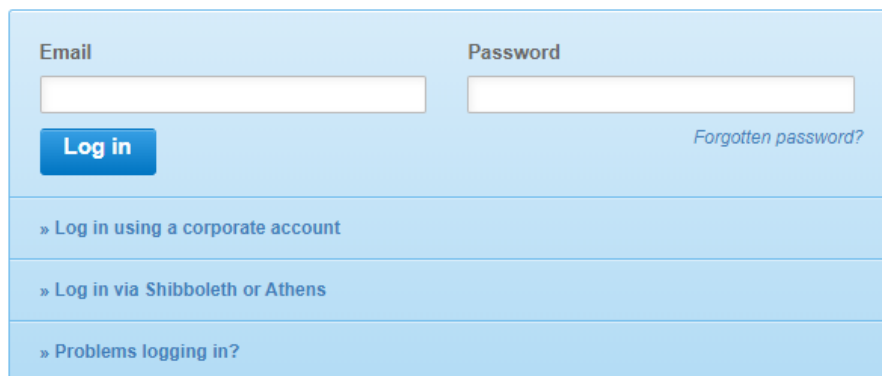Seleccione Sing up / Log in para tener acceso a todas las funciones que estén disponibles de Springer

Si ingresa el correo registrado y el password, se tendrá el acceso a su panel de la biblioteca
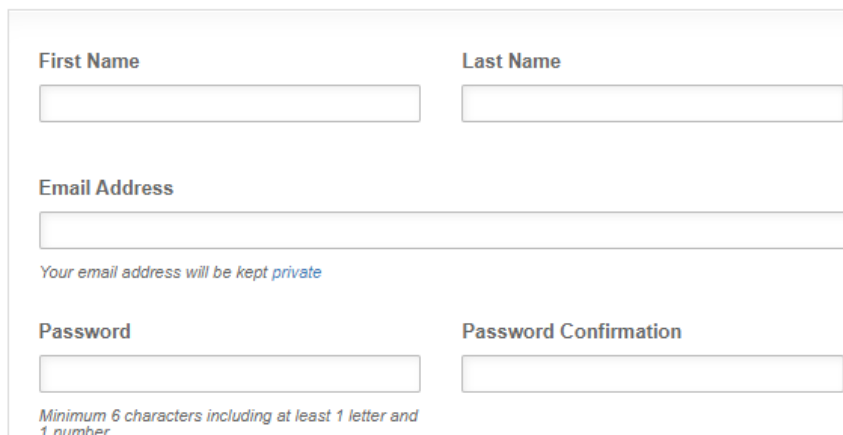
O debera de registrarse para obtener el acceso

Luego de loguearse tendrá acceso al panel

BUSQUEDA DE CONTENIDO

Puede realizar busquedas que incluyan preview o sean de pago, o pueden seleccionar los que si incluyan mas detalles en pdf o html

Seleccionamos un contenido

Ampliamos la información en el botón (Browse Volumes & Issues)

Acceso al Volume 1 – January 2020

Se podrá acceder al Download PDF o ver el articulo online

Archivo en PDF

ORIGINAL RESEARCH

SN

Multidimensional Feature Selection and High Performance ParalleX

A tool for detection of informative variables for big data

Karol Niedzielewski[1] · Maciej E. Marchwiany[1] · Radoslaw Piliszek[2] · Marek Michalewicz[1] · Witold Rudnicki[1,2,3]

Abstract

Great amount of stored information used in connection with Machine Learning and statistical methods enables high quality insight and analysis of data that leads to design of high precision predictive and classification systems. In the process of analysis, selection of most informative features is crucial for later quality of the designed system. In this report, we propose two implementations of multidimensional feature selection (MDFS) algorithm (Piliszek et al. in Mdfs-multidimensional feature selection. arXiv preprint. arXiv:1811.00631, 2018) that can be used in distributed environments for detection of all-relevant variables in data sets with discrete decision variable. While most methods discard information about interactions between features, MDFS is designed towards identification of informative variables that are not relevant when considered alone but are relevant in groups. We have developed software using C++ and High Performance ParalleX (HPX) (Kaiser et al. in STEllAR-GROUP/hpx: HPX V1.3.0: the C++ Standards library for parallelism and concurrency. 2019. https://doi.org/10.5281/zenodo.3189323, 2019) to achieve best performance, great scalability and portability. HPX is a library that uses lightweight threads, asynchronous communication, and asynchronous task submission based on the declarative criteria of work. These features enabled us to deeply explore granularity and parallelism of the MDFS algorithm. Software is prepared entirely in C++; therefore, calculations can be performed using CPUs on desktops, distributed systems, and any system with C++ compiler support. During testing on Cray XC40 (Okeanos) using artificially prepared data, we achieved 196 times acceleration on 256 nodes compared to a single node. From this point, ICM computing facility is capable of massively parallel feature engineering. The main purpose of the software is to enable researchers for more accurate genomics data analysis in search for multiple correlations in potential sources of the diseases.

Keywords Multidimensional feature selection · Mutual information · HPX · Distributed systems · Big data · Genomics

Introduction

Versión online

SN Computer Science
January 2020, 1:40 | Cite as

Multidimensional Feature Selection and High Performance ParalleX

A tool for detection of informative variables for big data

Authors   Authors and affiliations

Karol Niedzielewski ✉ , Maciej E. Marchwiany, Radoslaw Piliszek, Marek Michalewicz, Witold Rudnicki

Open Access   Original Research
First Online: 24 October 2019

62
Downloads

Part of the following topical collections:
- Modelling methods in Computer Systems, Networks and Bioinformatics

Download PDF ⬇

Cite article ▼

Article
Abstract
Introduction
Theory
Algorithm Implementation
Results
Conclusions
Future Work
Footnotes
Notes
References
Copyright information
About this article

Abstract

Great amount of stored information used in connection with Machine Learning and statistical methods enables high quality insight and analysis of data that leads to design of high precision predictive and classification systems. In the process of analysis, selection of most informative features is crucial for later quality of the designed system. In this report, we propose two