

REQUIREMENTS DOCUMENT

Batch Rename PDF files

By

Jayanth Anantharaman
Apoorva Vinod Gorur
Md Kamruzzaman Sarker

Wright State University, 2017

Table of Contents

1	Introduction	2
	1.1 Purpose of This Document	2
	1.2 Scope of the Product	2
	1.3 Definitions, acronyms and abbreviations	2
2	General Description	2
	2.1 Functional Requirements	3
	2.2 Performance Requirement	3
	2.3 Maintainability	3
	2.4 Platform/Environments	4
3	User Interface	4
4	Acceptance Test	4

Include
appendix?

1. Introduction

This document contains the requirements for Batch Rename PDF files system.

1.1. Purpose of this Document

This is intended to provide documentation of requirements for Batch Rename PDF files system.

1.2. Scope of the Product

The Scope of the product is to only rename the PDF files existing in a folder by parsing the PDF document to extract required details mentioned in one of the below sections. The product will not in any case edit or manipulate the contents of the PDF or delete any existing file.

1.3. Definitions, acronyms and abbreviations

Term	Definitions
Subject Classification	Broad category of the documents to be renamed.
Title	Title of the document. E.g.: Research paper title
Year	4-digit number representing the calendar year of the published document
Conference/Journal name	Venue of the document published
Author name	Name of the person(s) authored the document
Affiliation	Name of the institution/organization to which authors are affiliated

Table 1 - Definitions of terms

2. General Description

Batch Rename PDF files is a system which will rename the PDF files in a folder in batches with a specific format by extracting relevant information from within the PDF file.

2.1. Functional Requirements:

Was this decided upon by your team? What is the scenario when some of the information is not available?

May want to reword this.

1. The product must rename the PDF files within a directory by extracting one or many of the below mentioned information from within the PDF file according to the user choice.
 - a. The subject classification.
 - b. The title of the paper.
 - c. The year of publication.
 - d. Name of conference or journal of publication
 - e. List of author names (First Middle Last)
 - f. List of Institutions affiliated with the author.
2. Renaming should follow the below format.
`SUBJECT CLASSIFICATION_TITLE_YEAR_VENUE_AUTHORNAMES_AFFILIATIONS.pdf`
3. Product should report the number of files successfully renamed out of the given batch.
4. If any required information is missing in the pdf document, rename it with available information from the pdf document. If none of the required information are available, skip renaming the pdf document
5. If the pdf is password protected or not readable skip renaming the pdf document.
6. Files with .pdf extension should only be processed.
7. The product should not in any case edit or manipulate the contents of the PDF or delete any existing file

2.2. Performance Requirement:

1. Product should be able to rename a given batch of pdf files with 95% accuracy with reasonable efficiency.
2. Product should not crash or hang.
3. Proper error handling with error/log message to user.

2.3. Maintainability:

Source code should be documented and maintainable.

2.4. Platform/Environments:

Product should be available for Windows, Mac and Linux platforms.

3. User Interface

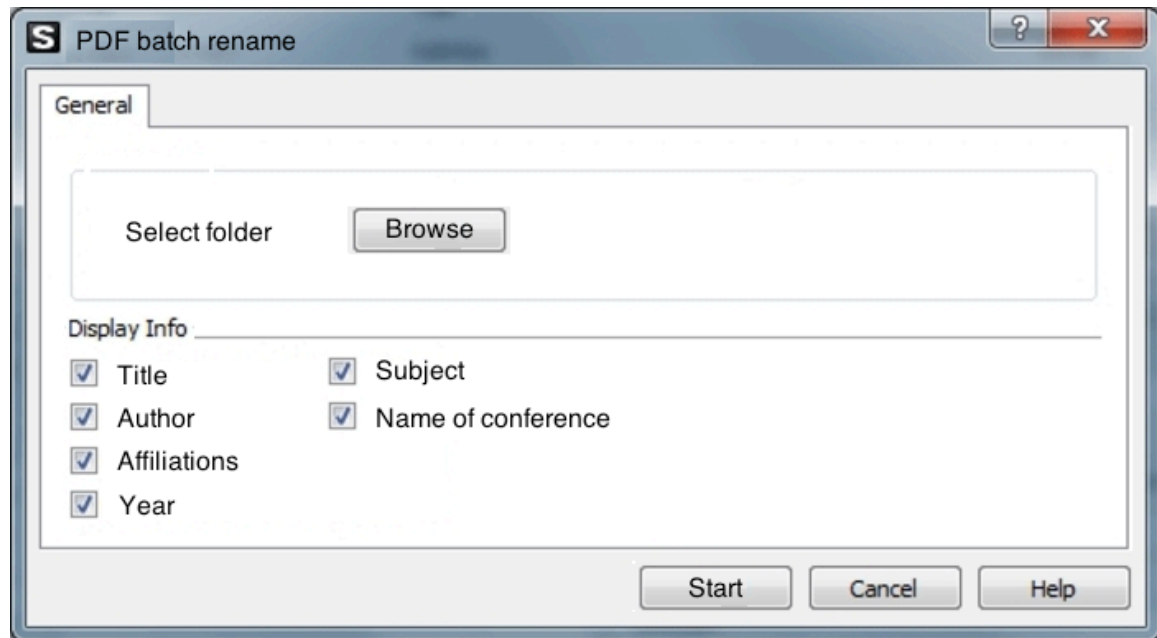


Figure 1: User Interface of PDF Batch Rename files

Figure 1 is a sample mock - up of the user interface to be developed in the product. Below are the functionalities expected from the user interface.

May be useful to include a text box that displays the folder the user selected. There is nothing to view what folder is about to be batch renamed.

1. It must include actionable buttons namely
 - a. Browse - Button to input the folder where documents are to be renamed
 - b. Checkboxes - Check box indicating the labels to be included while renaming the pdf documents
 - c. Start - Initiate the process of renaming the documents
 - d. Cancel - Gracefully stop the running process
 - e. Help - To provide user documentation of the product

4. Acceptance Test

Following are some sample snapshots of PDFs from which the product should be able to extract the required information to rename.

Complete set of documents are available in the following link¹

Communicated by Richard Lippmann

**Use of an Artificial Neural Network for Data Analysis in
Clinical Decision-Making: The Diagnosis of Acute
Coronary Occlusion**

William G. Baxt

*Department of Medicine, University of California,
San Diego Medical Center, San Diego, CA 92103 USA*

Figure 2: Research Paper Sample 1

Title: Use of an Artificial Neural Network for Data Analysis in Clinical Decision-Making
Author(s): William G. Baxt
Affiliation: University of California
Year: 1990
Subject: TBD
Name of Conference: NA

2014 IEEE Congress on Evolutionary Computation (CEC)
July 6-11, 2014, Beijing, China

**Evolving a Fuzzy Goal-Driven Strategy for the Game of Geister:
An Exercise in Teaching Computational Intelligence**

Andrew R. Buck, *Student Member IEEE*, Tanvi Banerjee, *Student Member IEEE*,
and James M. Keller, *Fellow IEEE*

Figure 3: Research Paper Sample 2

Title: Evolving a Fuzzy Goal-Driven Strategy for the Game of Geister

Author(s):

Andrew R Buck,
Tanvi Banerjee,
James M Keller

Affiliation: IEEE

Year: 2014

Subject: TBD

Name of Conference: IEEE Congress on Evolutionary Computation

The Collateral Damage of Internet Censorship by DNS Injection *

Sparks
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com

Neo[†]
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com

Tank
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com

Smith
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com

Dozer
Hovership Nebuchadnezzar
Zion Virtual Labs
zion.vlab@gmail.com

Figure 4: Research Paper Sample 3

Title: The Collateral Damage of Internet Censorship by DNS Injection

Author(s):

Spark,
Neo,
Tank,
Smith, Dozer

Affiliation: Zion Virtual Labs

Year: NA

Subject: TBD

Name of Conference: NA

Mobile Data Charging: New Attacks and Countermeasures

Chunyi Peng Chi-yu Li Guan-hua Tu Songwu Lu Lixia Zhang

Department of Computer Science, University of California, Los Angeles, CA 90095
{chunyp, lichiyu, ghtu, slu, lixia}@cs.ucla.edu

Figure 5: Research Paper Sample 4

Title: Mobile Data Charging: New Attacks and Countermeasures

Author(s):

Chunyi Peng,
Chi-yu Li,
Guan-hua Tu,
Songwu Lu,
Lixia Zhang

Affiliation: University of California

Year: NA

Subject: TBD

Name of Conference: NA

**“Foiling the Cracker”:
A Survey of, and Improvements to, Password Security[†]**

Daniel V. Klein

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15217
dvk@sei.cmu.edu
+1 412 268 7791

ABSTRACT

Figure 6: Research Paper Sample 5

Title: Foiling the Cracker
Author(s): Daniel V Klein
Affiliation: Carnegie Mellon University
Year: NA
Subject: TBD
Name of Conference: NA

QryGraph: A Graphical Tool for Big Data Analytics

Sanny Schmid, Ilias Gerostathopoulos, Christian Prehofer

Fakultät für Informatik

Technische Universität München

Munich, Germany

{schmidsa, gerostat, prehofer}@in.tum.de

Figure 7: Research Paper Sample 6

Title: QryGraph: A Graphical Tool for Big Data Analytics

Author(s):

Sanny Schmid,

Ilias Gerostathopoulos,

Christian Prehofer

Affiliation: Fakultät für Informatik Technische Universität München Munich, Germany

Year: NA

Subject: TBD

Name of Conference: NA

Rethinking High Performance Computing System Architecture for Scientific Big Data Applications

Yong Chen*, Chao Chen*, Yanlong Yin[†], Xian-He Sun[‡], Rajeev Thakur[‡], William D Gropp[§]

*Department of Computer Science, Texas Tech University, Email: yong.chen@ttu.edu, chao.chen@ttu.edu

[†]Department of Computer Science, Illinois Institute of Technology, Email: yyin2@ttu.edu, sun@ttu.edu

[‡]Mathematics and Computer Science Division, Argonne National Laboratory, Email: thakur@mcs.anl.gov

[§]Department of Computer Science, University of Illinois Urbana-Champaign, Email: wgropp@illinois.edu

Figure 8: Research Paper Sample 7

Title: Rethinking High Performance Computing System Architecture for Scientific Big Data Applications

Author(s):

Yong Chen,
Chao Chen,
Xian-He Sun Rajeev Thakur,
William D Gropp

Affiliation:

Texas Tech University, Illinois Institute of technology, Argonne National Laboratory,
University of Illinois Urbana-Champaign

Year: 2016

Subject: TBD

Name of Conference: IEEE TrustCom-BigDataSE-ISPA

A Named Data Network Approach to Energy Efficiency in IoT

Oliver Hahm
Inria

Emmanuel Baccelli
Inria

Thomas C. Schmidt
HAW

Matthias Wählisch
FU Berlin

Cédric Adjih
Inria

Figure 9: Research Paper Sample 8

Title: A Named Data Network Approach to Energy Efficiency in IoT

Author(s):

Oliver Hahm Inria,
Thomas C. Schmidt HAW,
Matthias Wählisch FU Berlin,
Cédric Adjih Inria

Affiliation: NA




Year: 2016

Subject: TBD


Name of Conference: NA

Research Paper

Immune Cell Repertoire and Their Mediators in Patients with Acute Myocardial Infarction or Stable Angina Pectoris

Wenwen Yan¹, Yanli Song², Lin Zhou¹, Jinfa Jiang¹, Fang Yang³, Qianglin Duan¹, Lin Che¹, Yuqin Shen¹, Haoming Song¹, Lemin Wang¹

1. Department of Cardiology, Tongji Hospital, Tongji University School of Medicine, Shanghai 200065, China;
2. Department of Emergency Medicine, Tongji Hospital, Tongji University School of Medicine, Shanghai 200065, China;
3. Department of Experimental Diagnosis, Tongji Hospital, Tongji University School of Medicine, Shanghai 200065, China.

 Corresponding authors: Lemin Wang, Haoming Song, Yuqin Shen, Department of Cardiology, Tongji Hospital, Tongji University School of Medicine, 389 Xincun Rd, Putuo District, Shanghai 200065, China; Tel: 86 21 66111329, Fax: 86 21 66111329, E-mail: wanglemin@tongji.edu.cn; songhao-ming@163.com; sy-1963@126.com.

© Ivyspring International Publisher. This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY-NC) license (<https://creativecommons.org/licenses/by-nc/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2016.08.05; Accepted: 2016.12.21; Published: 2017.02.08

Figure 10: Research Paper Sample 9

Title: Immune Cell Repertoire and Their Mediators in Patients with Acute Myocardial Infarction or Stable Angina Pectoris

Author(s):

Wenwen Yan,
Yanli Song,
Lin Zhou,
Jinfa Jiang,
Fang Yang,
Qianglin Duan,
Lin Che,
Yuqin Shen,
Haoming Song,
Lemin Wang

Affiliation: Tongji University School of Medicine, Shanghai

Year: 2017

Subject: TBD

Name of Conference: International Journal of Medical Sciences

Preordering using a Target-Language Parser via Cross-Language Syntactic Projection for Statistical Machine Translation

ISAO GOTO, National Institute of Information and Communications Technology, NHK,
and Kyoto University
MASAO UTIYAMA and EIICHIRO SUMITA, National Institute of Information
and Communications Technology
SADAO KUROHASHI, Kyoto University

Figure 11: Research Paper Sample 10

Title: Preordering using a Target-Language Parser via Cross-Language Syntactic Projection for Statistical Machine Translation

Author(s):

ISAO GOTO,
MASAO UTIYAMA,
EIICHIRO SUMITA,
SADAO KUROHASHI

Affiliation: National Institute of Information and Communications Technology, Kyoto University

Year: NA

Subject: TBD

Name of Conference: NA



Available online at www.sciencedirect.com



Decision Support Systems 42 (2006) 674–689

Decision Support
Systems

www.elsevier.com/locate/dsw

A new approach to classification based on association rule mining

Guoqing Chen^{*}, Hongyan Liu, Lan Yu, Qiang Wei, Xing Zhang

Department of Management Science and Engineering, School of Economics and Management, Tsinghua University, Beijing 100084, China

Received 19 February 2004; received in revised form 9 March 2005; accepted 9 March 2005

Available online 25 July 2005

Figure 12: Research Paper Sample 11

Subject Classification: Not available

Title: A new approach to classification based on association rule mining

Year: received-2004, accepted-2005

Conference/Journal name: Not available

Author name:

Guoqing Chen¹,

Hongyan Liu¹,

Lan Yu¹,

Qiang Wei¹,

Xing Zhang¹

Affiliation: Department of Management Science and Engineering, School of Economics and Management, Tsinghua University, Beijing 100084, China¹

Induction of Decision Trees

J.R. QUINLAN (munari!nswitgould.oz!quinlan@seismo.css.gov)
*Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney 2007,
Australia*

(Received August 1, 1985)

Key words: classification, induction, decision trees, information theory, knowledge acquisition, expert systems

Figure 13: Research Paper Sample 12

Subject Classification: classification, induction, decision trees, information theory, knowledge acquisition, expert systems

Title: Induction of Decision Trees

Year: 1986

Conference/Journal name: Kluwer Academic Publishers, Boston

Author name:

J.R. QUINLAN

Affiliation: Centre for Advanced Computing Sciences, New South Wales Institute of Technology, Sydney, Australia

XGBoost: A Scalable Tree Boosting System

Tianqi Chen
University of Washington
tqchen@cs.washington.edu

Carlos Guestrin
University of Washington
guestrin@cs.washington.edu

ABSTRACT

Tree boosting is a highly effective and widely used machine learning method. In this paper, we describe a scalable end-to-end tree boosting system called XGBoost, which is used widely by data scientists to achieve state-of-the-art results on many machine learning challenges. We propose a novel sparsity-aware algorithm for sparse data and weighted quantile sketch for approximate tree learning. More importantly, we provide insights on cache access patterns, data compression and sharding to build a scalable tree boosting system. By combining these insights, XGBoost scales beyond billions of examples using far fewer resources than existing systems.

Keywords

Large-scale Machine Learning

problems. Besides being used as a stand-alone predictor, it is also incorporated into real-world production pipelines for ad click through rate prediction [15]. Finally, it is the de-facto choice of ensemble method and is used in challenges such as the Netflix prize [3].

In this paper, we describe XGBoost, a scalable machine learning system for tree boosting. The system is available as an open source package². The impact of the system has been widely recognized in a number of machine learning and data mining challenges. Take the challenges hosted by the machine learning competition site Kaggle for example. Among the 29 challenge winning solutions³ published at Kaggle's blog during 2015, 17 solutions used XGBoost. Among these solutions, eight solely used XGBoost to train the model, while most others combined XGBoost with neural nets in ensembles. For comparison, the second most popular method, deep neural nets, was used in 11 solutions. The success

Figure 14: Research Paper Sample 13

Subject Classification: Large-scale Machine Learning

Title: XGBoost: A Scalable Tree Boosting System

Year: Not Available

Conference/Journal name: Not Available

Author name:

Carlos Guestrin¹,
Carlos Guestrin¹

Affiliation: University of Washington¹

Obesity and Other Cancers

Lin Yang, Bettina F. Drake, and Graham A. Colditz

Lin Yang, Bettina F. Drake, and Graham A. Colditz, Washington University School of Medicine and Siteman Cancer Center, St Louis, MO; and Lin Yang, Center for Public Health, Medical University of Vienna, Vienna, Austria.

Published online ahead of print at www.jco.org on November 7, 2016.

Authors' disclosures of potential conflicts of interest and author contributions are found at the end of this article.

Corresponding author: Graham A. Colditz, MD, DrPH, Washington University School of Medicine, 660 S Euclid Ave, Campus Box 8100, St Louis, MO 63110; e-mail: colditzg@wustl.edu.

© 2016 by American Society of Clinical Oncology

0732-183X/16/3435w-4231w/\$20.00

DOI: 10.1200/JCO.2016.68.4837

A B S T R A C T

Purpose

Evidence on overweight, obesity, and an increased risk of cancer continues to accumulate and was updated in the 2016 handbook on weight control from the International Agency for Research on Cancer (IARC). The underlying primary data, together with dose-response meta-analysis and, finally, pooled analysis of individual participant data, add insight into the relation between obesity and cancer risk and prognosis. We summarize the evidence for mortality from prostate cancer, hematologic malignancies, and kidney cancer.

Methods

We reviewed pooled analysis of rare end points across cohorts, regardless of primary results reported from the individual studies, further reducing risk of publication bias. Of these cancer sites, only kidney cancer was included in the IARC 2002 report, although mortality from prostate cancer and hematologic malignancies was noted in the American Cancer Society prospective cohort study in 2003. The 2016 update from the IARC added details for prostate and hematologic malignancies, classifying the evidence as sufficient to conclude that avoiding excess body fatness lowers the risk of multiple myeloma but found that the evidence for it lowering the risk of prostate cancer mortality or diffuse large B-cell lymphoma was limited.

Results

A higher body mass index is associated with an increased risk of advanced prostate cancer and

Figure 15: Research Paper Sample 14

Subject Classification: Not Available

Title: Obesity and Other Cancers

Year: 2016

Conference/Journal name: JOURNAL OF CLINICAL ONCOLOGY

Author name:

Lin Yang^{1,2},
Bettina F. Drake¹, and
Graham A. Colditz¹

Affiliation:

Washington University School of Medicine and Siteman Cancer Center, St Louis, MO¹.

Center for Public Health, Medical University of Vienna, Vienna, Austria².

Experimental test of Landauer's principle in single-bit operations on nanomagnetic memory bits

Jeongmin Hong,¹ Brian Lambson,² Scott Dhuey,³ Jeffrey Bokor^{1*}

2016 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC). 10.1126/sciadv.1501492

Minimizing energy dissipation has emerged as the key challenge in continuing to scale the performance of digital computers. The question of whether there exists a fundamental lower limit to the energy required for digital operations is therefore of great interest. A well-known theoretical result put forward by Landauer states that any irreversible single-bit operation on a physical memory element in contact with a heat bath at a temperature T requires at least $k_B T \ln(2)$ of heat be dissipated from the memory into the environment, where k_B is the Boltzmann constant. We report an experimental investigation of the intrinsic energy loss of an adiabatic single-bit reset operation using nanoscale magnetic memory bits, by far the most ubiquitous digital storage technology in use today. Through sensitive, high-precision magnetometry measurements, we observed that the amount of dissipated energy in this process is consistent (within 2 SDs of experimental uncertainty) with the Landauer limit. This result reinforces the connection between "information thermodynamics" and physical systems and also provides a foundation for the development of practical information processing technologies that approach the fundamental limit of energy dissipation. The significance of the result includes insightful direction for future development of information technology.

Figure 16: Research Paper Sample 15

Subject Classification: Not Available

Title: Experimental test of Landauer's principle in single-bit operations on nanomagnetic memory bits

Year: Not Available

Conference/Journal name: Not Available

Author names:

Jeongmin Hong¹,
Brian Lambson²,
Scott Dhuey³,
Jeffrey Bokor¹

Affiliation:

Electrical Engineering and Computer Sciences, University of California, Berkeley, Berkeley, CA 94720, USA¹.

Haynes and Boone LLP, 525 University Avenue, Palo Alto, CA 94301, USA².

The Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA 94720, USA³.

MAY WANT TO ADD REFERENCES TO EACH PDF LISTED.

Appendix

[1]<https://drive.google.com/drive/folders/0BwYt86HaSAW9RG1NdGMtTGYzdkE?usp=sharing>