

Institutional Sign In

Browse

My Settings

Get Help

Subscribe

Advertisement

Browse Conferences > Cyber and IT Service Manageme...

Incremental technique with set of frequent word item sets for mining large Indonesian text data

Related Articles

Improving smart card security using self-timed circuits

Efficient key-frame extraction and video analysis

[View All](#)

Sign In or Purchase to View Full Text

20 Full Text Views

4 Author(s)

Dian Sa'adillah Maylawati ; Muhammad Ali Ramdhani ; Ali Rahman ; Wahyudin Darmalaksana

[View All Authors](#)

Abstract	Authors	Figures	References	Citations	Keywords	Metrics	Media
-----------------	---------	---------	------------	-----------	----------	---------	-------

Abstract:

Indonesian text data from social media is one of large text data that interesting to be mined. Mining the insight knowledge from large text data need more effort and time to processed. Moreover, Indonesian text data from social media contains natural language, including slang that require special treatment. We propose incremental technique for more efficient mining process of large text data with Set of Frequent Word Itemset (SFWI) representation that had been proven capable to keep the meaning of Indonesian text well. We compared Frequent Pattern Growth (FP-Growth) algorithm for not incremental mining and Compact Pattern Growth (CP-Tree) algorithm for incremental mining. The result of experiment with 3,200, 5,000, 110,000, and 239,496 text data form Twitter showed that the incremental technique capable to reduce time process and memory usage for mining Indonesian large text data. Incremental technique with CP-Tree could decrease time process and memory usage so that time process was about 1.66 times faster and 1.84 times more efficient for memory usage than with FP-Growth which was not incremental.

Published in: Cyber and IT Service Management (CITSM), 2017 5th International Conference on

Date of Conference: 8-10 Aug. 2017

INSPEC Accession Number: 17316221

Date Added to IEEE Xplore: 30 October 2017

DOI: 10.1109/CITSM.2017.8089224

ISBN Information:

Publisher: IEEE

Conference Location: Denpasar, Indonesia

Advertisement

Contents

Download PDF

Download Citation

View References

Email

Print

I. Introduction

Sustainable development involves natural and human resource optimizing process [1], as well as technology. Social media as a technology is strongly influenced human resources especially its users. Social media still be an interesting object for mining research. Many information can be extracted from social media. Text data is one of data form in social media that contain the meaning of ideas and usually written with natural language, even more contain slangs. Therefore, there are several phase and special treatment to process text data from social media. All of text data processing need effort and time. The problem is text data from social media is huge. In one minute Facebook stores more than 3.4 million post, more than 350,000 tweets from Twitter, and currently 2.5 quantillion data flow on social every day, including text data [2]. Besides that, Indonesia is one of the largest social media user which gives a lot of data in social media [3] [4]. Based on the facts above, reducing effort and time for mining large text data is an important things.

Full Text

Abstract

Authors

Figures

Request Permissions

Export to Collabratec

Alerts

Read document

References

Keywords

Citations

IEEE Keywords

Keywords

Data mining, Media, Social network services, TV, Feature extraction, Natural languages, Databases

Back to Top

INSPEC: Controlled Indexing

data mining, data structures, natural language processing, social networking (online)

INSPEC: Non-Controlled Indexing

Indonesian text data, social media, incremental mining, data mining, frequent pattern growth algorithm, frequent word itemset representation, natural language, SFWI, compact pattern growth algorithm, Twitter, insight knowledge mining

Author Keywords

Incremental mining, Set of FWI, FWI, frequent pattern, text mining, FP-Growth, CP-Tree

Authors

Dian Sa'adillah Maylawati
Department of Informatics, UIN Sunan Gunung Djati Bandung

Muhammad Ali Ramdhani
Department of Informatics, UIN Sunan Gunung Djati Bandung

Ali Rahman
Department of Informatics, UIN Sunan Gunung Djati Bandung

Wahyudin Darmalaksana
Research Center, UIN Sunan Gunung Djati Bandung

Related Articles

Improving smart card security using self-timed circuits
S. Moore; R. Anderson; P. Cunningham; R. Mullins; G. Taylor

Efficient key-frame extraction and video analysis
J. Calic; E. Izquierdo

On Arabic-English cross-language information retrieval: a machine translation approach
M. Aljlayl; O. Frieder; D. Grossman

Automatic query expansion based on directed divergence
D. Cai; C.J. van Rijsbergen

Dependable Semantic Web
B. Thuraisingham; E. Hughes; D. Allen

A method for compressing lexicons
S. Ristov; E. Laporte

Diagnostically lossless compression of pipeline inspection data
Wei-Ching Tham; S.I. Woolley; S. Cribbs; D. Anderson

Semi-discrete matrix transforms (SDD) for image and video compression
S. Zyto; A. Grama; W. Szpankowski

Object testing in ITEE
S. Ohara; F. Tsunoda; H. Maezawa; M. Hui; T. Wang; P.C.Y. Sheu; R. Paul

Maintaining consistency of the security policy using timestamp ordering
S. Ngamsuriyaroj; T.F. Keefe; A.R. Hurson

IEEE Account

- » Change Username/Password
- » Update Address

Purchase Details

- » Payment Options
- » Order History
- » View Purchased Documents

Profile Information

- » Communications Preferences
- » Profession and Education
- » Technical Interests

Need Help?

- » **US & Canada:** +1 800 678 4333
- » **Worldwide:** +1 732 981 0060
- » Contact & Support

[About IEEE Xplore](#) | [Contact Us](#) | [Help](#) | [Accessibility](#) | [Terms of Use](#) | [Nondiscrimination Policy](#) | [Sitemap](#) | [Privacy & Opting Out of Cookies](#)

A not-for-profit organization, IEEE is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity.
© Copyright 2018 IEEE - All rights reserved. Use of this web site signifies your agreement to the terms and conditions.