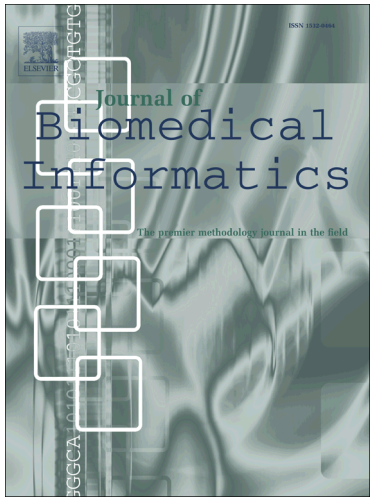
Accepted Manuscript



PDF text classification to leverage information extraction from publication reports

Duy Duc An Bui, Guilherme Del Fiol, Siddhartha Jonnalagadda

PII: S1532-0464(16)30017-X
DOI: http://dx.doi.org/10.1016/j.jbi.2016.03.026
Reference: YJBIN 2556

To appear in: *Journal of Biomedical Informatics*

Received Date: 23 September 2015
Revised Date: 22 March 2016
Accepted Date: 31 March 2016

Please cite this article as: Bui, D.D.A., Fiol, G.D., Jonnalagadda, S., PDF text classification to leverage information extraction from publication reports, *Journal of Biomedical Informatics* (2016), doi: http://dx.doi.org/10.1016/j.jbi.2016.03.026

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

PDF text classification to leverage information extraction from publication reports

Duy Duc An Bui, BS, PhD[1,2]
Guilherme Del Fiol, MD, PhD[1]
Siddhartha Jonnalagadda, PhD[2]

[1]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, USA
[2]Department of Preventive Medicine-Health and Biomedical Informatics, Northwestern University, Chicago, IL, USA

Corresponding Author:
Duy Duc An Bui, PhD
University of Utah
Biomedical Informatics
421 Wakara Way, Ste 140
Salt Lake City, UT 84108-3514
Phone: (801) 581-4080 / Fax: (801) 581-4297 (Fax)
duy.bui@utah.edu or bdaduy@gmail.com

Highlights

- Full-text publications in PDF format are the main sources used in the development of systematic reviews.
- PDF reports used in the development of Cochrane systematic reviews were used as gold a standard.
- The multi-pass sieve algorithm is a powerful alternative approach to the machine learning approaches in the classification of PDF texts.
- The PDF text classification algorithm improved performance of an information extraction system that identifies clinical outcome mentions from full-text publications.

Keywords: Text Classification, Natural Language Processing, Document Analysis, Machine Learning

Download English Version:

https://daneshyari.com/en/article/6927782

Download Persian Version:

https://daneshyari.com/article/6927782

Daneshyari.com