Accepted Manuscript

Title: Text mining of cancer-related information: Review of current status and future directions

Author: Irena Spasić Jacqueline Livsey John A. Keane Goran Nenadić

PII: S1386-5056(14)00110-5
DOI: http://dx.doi.org/doi:10.1016/j.ijmedinf.2014.06.009
Reference: IJB 3104

To appear in: *International Journal of Medical Informatics*

Received date: 23-10-2013
Revised date: 12-6-2014
Accepted date: 14-6-2014

Please cite this article as: I. Spasić, J. Livsey, J.A. Keane, G. Nenadić, Text mining of cancer-related information: review of current status and future directions, *International Journal of Medical Informatics* (2014), http://dx.doi.org/10.1016/j.ijmedinf.2014.06.009

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

Text mining of cancer-related information:

review of current status and future directions

Irena Spasić[a]*, Jacqueline Livsey[b], John A. Keane[c], Goran Nenadić[c]

[a]*School of Computer Science & Informatics, Cardiff University, Cardiff, CF24 3AA, UK*

[b]*Clinical Outcomes Unit, The Christie NHS Foundation Trust, Manchester, M20 4BX, UK*

[c]*School of Computer Science, The University of Manchester, Manchester, M13 9PL, UK*

*Corresponding author. E-mail address: i.spasic@cs.cardiff.ac.uk. Telephone number: +44(0)29 2087 0320

Abstract

Purpose: This paper reviews the research literature on text mining (TM) with the aim to find out (1) which cancer domains have been the subject of TM efforts, (2) which knowledge resources can support TM of cancer-related information and (3) to what extent systems that rely on knowledge and computational methods can convert text data into useful clinical information. These questions were used to determine the current state of the art in this particular strand of TM and suggest future directions in TM development to support cancer research.

Methods: A review of the research on TM of cancer-related information was carried out. A literature search was conducted on the Medline database as well as IEEE Xplore and ACM digital libraries to address the interdisciplinary nature of such research. The search results were supplemented with the literature identified through Google Scholar.

Results: A range of studies have proven the feasibility of TM for extracting structured information from clinical narratives such as those found in pathology or radiology reports. In this article, we provide a critical overview of the current state of the art for TM related to cancer. The review highlighted a strong bias towards symbolic methods, e.g. named entity recognition (NER) based on dictionary lookup and information extraction (IE) relying on pattern matching. The F-measure of NER ranges between 80% and 90%, while that of IE for simple tasks is in the high 90s. To further improve

Accepted Manuscript

Download English Version:

https://daneshyari.com/en/article/6926958

Download Persian Version:

https://daneshyari.com/article/6926958

Daneshyari.com