

## Accepted Manuscript

Tumor reference resolution and characteristic extraction in radiology reports for liver cancer stage prediction

Wen-wai Yim, Sharon W. Kwan, Meliha Yetisgen

PII: S1532-0464(16)30138-1

DOI: <http://dx.doi.org/10.1016/j.jbi.2016.10.005>

Reference: YJBIN 2659

To appear in: *Journal of Biomedical Informatics*

Received Date: 27 May 2016

Accepted Date: 7 October 2016



Please cite this article as: Yim, W-w., Kwan, S.W., Yetisgen, M., Tumor reference resolution and characteristic extraction in radiology reports for liver cancer stage prediction, *Journal of Biomedical Informatics* (2016), doi: <http://dx.doi.org/10.1016/j.jbi.2016.10.005>

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.

# Tumor reference resolution and characteristic extraction in radiology reports for liver cancer stage prediction

Wen-wai Yim<sup>a</sup>, Sharon W. Kwan<sup>b</sup>, Meliha Yetisgen<sup>a,c,\*</sup>

<sup>a</sup>Biomedical and Health Informatics, University of Washington

<sup>b</sup>Radiology, University of Washington

<sup>c</sup>Linguistics, University of Washington

---

## Abstract

**Background:** Anaphoric references occur ubiquitously in clinical narrative text. However, the problem, still very much an open challenge, is typically less aggressively focused on in clinical text domain applications. Furthermore, existing research on reference resolution is often conducted disjointly from real-world motivating tasks.

**Objective:** In this paper, we present our machine-learning system that automatically performs reference resolution and a rule-based system to extract tumor characteristics, with component-based and end-to-end evaluations. Specifically, our goal was to build an algorithm that takes in tumor templates and outputs tumor characteristic, e.g. tumor number and largest tumor sizes, necessary for identifying patient liver cancer stage phenotypes.

**Results:** Our reference resolution system reached a modest performance of 0.66 F1 for the averaged MUC, B-cubed, and CEAF scores for coreference resolution and 0.43 F1 for particularization relations. However, even this modest performance was helpful to increase the automatic tumor characteristics annotation substantially over no reference resolution.

**Conclusion:** Experiments revealed the benefit of reference resolution even for relatively simple tumor characteristics variables such as largest tumor size. However we found that different overall variables had different tolerances to reference resolution upstream errors, highlighting the need to characterize systems by end-to-end evaluations.

**Keywords:** Natural language processing, Information extraction, Reference resolution, Radiology report, Cancer stages, Liver cancer

---

## 1. Introduction

*Reference resolution* is the task of identifying expressions in text that refer to the same real-world entities. In natural discourse, humans readily em-

---

\*Corresponding author at Biomedical and Health Informatics, Department of Biomedical Informatics and Medical Education, University of Washington, Box 358047, Seattle, WA 98195

Email address: melihay@uw.edu (Meliha Yetisgen)

Download English Version:

<https://daneshyari.com/en/article/4966948>

Download Persian Version:

<https://daneshyari.com/article/4966948>

[Daneshyari.com](https://daneshyari.com)