

Using caTIES as a case-finding tool in tissue repositories: system challenges and lessons learned

Micaiah H. Evans, MD, MPH, B. W. Trevor Rohm, MD, MS, Frederick A. Schultz, MA, and Philip J. Kroth, MD, MS

The Cancer Text Information Extraction System (caTIES) comprises a set of tools that can dramatically streamline the tissue sample identification process using full text surgical pathology reports (SRPs) and natural language processing functionality. CaTIES also has the capability to de-identify SRPs. These functionalities can be of great value to tissue repositories by reducing human reviewer costs, making tissue with uncoded text-based SPRs much more available for research. Because caTIES is composed of multiple components, and has exact specifications for particular versions and configurations of these components, the installation and setup of caTIES presents a number of unique challenges. In addition, finding software support is often problematic as much of the documentation needed is incomplete, out of date, or located in obscure places. We summarized our experiences with setting up and configuring an instance of caTIES at the University of New Mexico's Tissue Repository and Experimental Pathology Laboratory by categorizing these challenges into four categories: 1) the pre-installation process; 2) the caTIES installation process; 3) program use and the user's manual; and 4) help and support. We describe the methods that we employed and the resources that we utilized in our successful installation. To evaluate the efficacy of caTIES for extracting accurate and relevant information from text-based SPRs, we are performing a comparison of the accuracy caTIES to that of manual human review.



This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.