

Resource Summary Report

Generated by PRECISE-TBI on Aug 3, 2026

caTIES - Cancer Text Information Extraction System

RRID:SCR_003444

Type: Tool

Proper Citation

caTIES - Cancer Text Information Extraction System (RRID:SCR_003444)

Resource Information

URL: <http://caties.cabig.upmc.edu/>

Proper Citation: caTIES - Cancer Text Information Extraction System (RRID:SCR_003444)

Description: The Cancer Text Information Extraction System (caTIES) provides tools for de-identification and automated coding of free-text structured pathology reports. It also has a client that can be used to search these coded reports. The client also supports Tissue Banking and Honest Broker operations. caTIES focuses on two important challenges of bioinformatics * Information extraction (IE) from free text * Access to tissue. Regarding the first challenge, information from free-text pathology documents represents a vital and often underutilized source of data for cancer researchers. Typically, extracting useful data from these documents is a slow and laborious manual process requiring significant domain expertise. Application of automated methods for IE provides a method for radically increasing the speed and scope with which this data can be accessed. Regarding the second challenge, there is a pressing need in the cancer research community to gain access to tissue specific to certain experimental criteria. Presently, there are vast quantities of frozen tissue and paraffin embedded tissue throughout the country, due to lack of annotation or lack of access to annotation these tissues are often unavailable to individual researchers. caTIES has three goals designed to solve these problems: * Extract coded information from free text Surgical Pathology Reports (SPRs), using controlled terminologies to populate caBIG-compliant data structures. * Provide researchers with the ability to query, browse and create orders for annotated tissue data and physical material across a network of federated sources. With caTIES the SPR acts as a locator to tissue resources. * Pioneer research for distributed text information extraction within the context of caBIG. caTIES focuses on IE from SPRs because they represent a high-dividend target for automated analysis. There are millions of SPRs in each major hospital system, and SPRs contain important information for researchers. SPRs act as tissue locators by indicating the presence of tissue blocks, frozen tissue and other resources, and by identifying the relationship of the tissue block to significant landmarks such as tumor margins. At present, nearly all important data within SPRs are embedded within loosely-structured free-text. For these reasons, SPRs were chosen to be

coded through caTIES because facilitating access to information contained in SPRs will have a powerful impact on cancer research. Once SPR information has been run through the caTIES Pipeline, the data may be queried and inspected by the researcher. The goal of this search may be to extract and analyze data or to acquire slides of tissue for further study. caTIES provides two query interfaces, a simple query dashboard and an advanced diagram query builder. Both of these interfaces are capable of NCI Metathesaurus, concept-based searching as well as string searching. Additionally, the diagram interface is capable of advanced searching functionalities. An important aspect of the interface is the ability to manage queries and case sets. Users are able to vet query results and save them to case sets which can then be edited at a later time. These can be submitted as tissue orders or used to derive data extracts. Queries can also be saved, and modified at a later time. caTIES provides the following web services by default: MMTx Service, TIES Coder Service

Abbreviations: caTIES

Synonyms: Cancer Text Information Extraction System, Cancer Text Information Extraction System (caTIES)

Resource Type: web service, software resource, data access protocol, data processing software, software application

Defining Citation: [PMID:20442142](https://pubmed.ncbi.nlm.nih.gov/20442142/)

Keywords: extraction, cancer, code, de-identification, information, paraffin, pathology, research, structure, surgical, system, tissue, tool, text, natural language processing, tissue banking, translational research, data sharing, collaboration, natural language processing, text-processing, text-mining, grid computing, service oriented architecture, query visualization, medical record, bioinformatics, automated coding

Funding: Cancer Biomedical Informatics Grid contract 79207CBS10;
NCI R01 CA132672;
NCI U01 CA 091343;
NCRR U54 RR023506-01

Availability: Open unspecified license

Resource Name: caTIES - Cancer Text Information Extraction System

Resource ID: SCR_003444

Alternate IDs: nif-0000-33212

Record Creation Time: 20220129T080219+0000

Record Last Update: 20260803T040341+0000

Ratings and Alerts

No rating or validation information has been found for caTIES - Cancer Text Information Extraction System.

No alerts have been found for caTIES - Cancer Text Information Extraction System.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We have not found any literature mentions for this resource.