

Resource Summary Report

Generated by SPARC SAWG on Aug 9, 2026

cTAKES

RRID:SCR_006379

Type: Tool

Proper Citation

cTAKES (RRID:SCR_006379)

Resource Information

URL: <http://ctakes.apache.org>

Proper Citation: cTAKES (RRID:SCR_006379)

Description: An open-source natural language processing system for information extraction from electronic medical record clinical free-text. This is a system through which one creates one or more pipelines to process clinical notes and to identify clinical named entities. It processes clinical notes, identifying types of clinical named entities, drugs, diseases/disorders, signs/symptoms, anatomical sites and procedures. Each named entity that is found is given attributes for the text span, the ontology mapping code, the context (family history of, current, unrelated to patient), and negated/not negated. cTAKES is built on the UIMA framework. cTAKES 2.5 does not provide a GUI of its own for installation or processing. The cTAKES documentation shows how to use the GUIs provided by the UIMA framework, and how to run cTAKES from a command line. Before using cTAKES you need to know that cTAKES does not provide any mechanisms of its own to handle patient data securely. It is assumed that cTAKES is installed on a system that can process patient data, or that any data being processed by cTAKES has already been through a deidentification step in order to comply with any applicable laws. The tool has been developed and deployed at Mayo Clinic since early 2000.

Abbreviations: cTAKES

Synonyms: cTAKES - clinical Text Analytics and Knowledge Extraction System, Clinical Text Analysis and Knowledge Extraction System

Resource Type: text-mining software, software resource, software application, source code

Defining Citation: [PMID:23286462](#), [PMID:20819853](#)

Keywords: natural language processing, information extraction, electronic medical record, medical record, clinical, free-text, annotation, unstructured information management architecture, uima

Funding: IBM UIMA ;
SHARPN Strategic Health IT Advanced Research Projects Area 4: Secondary Use of EHR Data Cooperative Agreement from the HHS Office of the National Coordinator Washington DC DHHS 90TR000201

Availability: Open-source

Resource Name: cTAKES

Resource ID: SCR_006379

Alternate IDs: nlx_152159

Old URLs: <https://wiki.nci.nih.gov/display/VKC/cTAKES+2.5>

Record Creation Time: 20220129T080235+0000

Record Last Update: 20260808T185858+0000

Ratings and Alerts

No rating or validation information has been found for cTAKES.

No alerts have been found for cTAKES.

Data and Source Information

Source: [SciCrunch Registry](#)

Usage and Citation Metrics

We found 45 mentions in open access literature.

Listed below are recent publications. The full list is available at [SPARC SAWG](#) .

Abdullah , et al. (2026) Explainable Patient-Level Cognitive Impairment Screening via Temporal, Semantic, and Psycholinguistic Multimodal AI. Journal of Intelligence, 14(4).

Abedi V, et al. (2024) Machine Learning-Based Prediction of Stroke in Emergency Departments. Therapeutic advances in neurological disorders, 17, 17562864241239108.

McMurry AJ, et al. (2024) Moving Biosurveillance Beyond Coded Data Using AI for Symptom Detection From Physician Notes: Retrospective Cohort Study. Journal of medical Internet research, 26, e53367.

Abu-Salih B, et al. (2023) Healthcare knowledge graph construction: A systematic review of the state-of-the-art, open issues, and opportunities. Journal of big data, 10(1), 81.

Afshar M, et al. (2023) Deployment of Real-time Natural Language Processing and Deep Learning Clinical Decision Support in the Electronic Health Record: Pipeline Implementation for an Opioid Misuse Screener in Hospitalized Adults. JMIR medical

informatics, 11, e44977.

Miller TA, et al. (2023) The SMART Text2FHIR Pipeline. medRxiv : the preprint server for health sciences.

Botlagunta M, et al. (2023) Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. Scientific reports, 13(1), 485.

Martin EA, et al. (2023) Hypertension identification using inpatient clinical notes from electronic medical records: an explainable, data-driven algorithm study. CMAJ open, 11(1), E131.

Kumar S, et al. (2022) Patient Representation Learning From Heterogeneous Data Sources and Knowledge Graphs Using Deep Collective Matrix Factorization: Evaluation Study. JMIR medical informatics, 10(1), e28842.

Lin Y, et al. (2022) External validation of a machine learning classifier to identify unhealthy alcohol use in hospitalized patients. Addiction (Abingdon, England), 117(4), 925.

Ganoe CH, et al. (2021) Natural language processing for automated annotation of medication mentions in primary care visit conversations. JAMIA open, 4(3), ooab071.

Zong N, et al. (2021) Leveraging Genetic Reports and Electronic Health Records for the Prediction of Primary Cancers: Algorithm Development and Validation Study. JMIR medical informatics, 9(5), e23586.

Afshar M, et al. (2021) External validation of an opioid misuse machine learning classifier in hospitalized adult patients. Addiction science & clinical practice, 16(1), 19.

Sun Y, et al. (2021) Building an OMOP common data model-compliant annotated corpus for COVID-19 clinical trials. Journal of biomedical informatics, 118, 103790.

Cooley-Rieders K, et al. (2021) Physician documentation matters. Using natural language processing to predict mortality in sepsis. Intelligence-based medicine, 5.

Dey V, et al. (2021) A Pipeline to Understand Emerging Illness Via Social Media Data Analysis: Case Study on Breast Implant Illness. JMIR medical informatics, 9(11), e29768.

Kulchak Rahm A, et al. (2021) User testing of a diagnostic decision support system with machine-assisted chart review to facilitate clinical genomic diagnosis. BMJ health & care informatics, 28(1).

Tsuji S, et al. (2021) Developing a RadLex-Based Named Entity Recognition Tool for Mining Textual Radiology Reports: Development and Performance Evaluation Study. Journal of medical Internet research, 23(10), e25378.

Miller TA, et al. (2020) Experiences implementing scalable, containerized, cloud-based NLP for extracting biobank participant phenotypes at scale. JAMIA open, 3(2), 185.

Pontikos N, et al. (2020) Phenogenon: Gene to phenotype associations for rare genetic diseases. PloS one, 15(4), e0230587.