SAP Connected Health Platform
Document Version: 1.0 – 2016-12-13

SAP Connected Health Platform Text Analysis
Language Reference
# Content

1. **SAP Connected Health Platform Text Analysis Language Reference.** ........................................... 3

2. **About this Guide.** ......................................................................................................................... 4
   2.1 Intended Audience. ...................................................................................................................... 4

3. **Functionality Overview.** ............................................................................................................. 5
   3.1 Overview of Clinical Entity Extraction. ....................................................................................... 6
   3.2 Overview of Clinical Fact Extraction. .......................................................................................... 6

4. **EXTRACTION_CORE_MEDICAL.** ............................................................................................ 7
   4.1 Clinical Entity Extraction. ............................................................................................................ 8
      - med_MED/RxNorm. .................................................................................................................. 9
      - med_DIS/NCIm. ...................................................................................................................... 9
      - med_SYMP/NCIm. .................................................................................................................. 10
      - med_LAB_TEST/NCIm. .......................................................................................................... 11
      - med_PROC_D/NCIm. ............................................................................................................. 11
      - med_PROC_T/NCIm. ............................................................................................................. 12
      - med_ANAT/NCIm. .................................................................................................................. 13
   4.2 Medical Section Headings Extraction. ....................................................................................... 13
   4.3 Clinical Entity Assertions. .......................................................................................................... 18
      - Modality Assertion. .............................................................................................................. 19
      - Polarity Assertion. ................................................................................................................ 20
      - Reference Assertion. ............................................................................................................ 20
   4.4 Clinical Entity Relations. .......................................................................................................... 21

5. **EXTRACTION_CORE_PATHOLOGY.** ......................................................................................... 22
   5.1 Pathology Entity Extraction. .................................................................................................... 23
      - med_HISTO/ICDO3. .............................................................................................................. 23
      - med_GENE/NCBI. ................................................................................................................. 24
   5.2 Medical Section Headings Extraction for Pathology. ................................................................ 24
   5.3 Pathology Entity Assertions. .................................................................................................... 24
   5.4 Pathology Entity Relations. ..................................................................................................... 25
1 SAP Connected Health Platform Text Analysis Language Reference
2  About this Guide

This guide focuses on the specific text-analysis features for the health domain. For more general information about the core text analysis features delivered with the SAP HANA platform, such as linguistic analysis, core extraction, and sentiment analysis fact extraction, see the SAP HANA Text Analysis Language Reference Guide.

This guide contains the following information:

- Overview and conceptual information about the healthcare-specific extraction features provided by SAP Connected Health.
- A reference section describing the behavior of the supported language module during extraction.

Related Information

Intended Audience [page 4]

2.1  Intended Audience

This guide is intended for users who want to customize and enhance the text analysis processes and, therefore, need a deeper understanding of the related concepts.

This document assumes the following:

- You are an application developer or consultant working on enhancing the text analysis and extraction processes.
- You understand the specific requirements of your organization with regards to text extraction.
Functionality Overview

Text analysis in SAP HANA is based on a suite of natural-language processing capabilities that model and structure the content of texts, for example, medical records. This advanced text processing capability forms the foundation for a variety of applications such as search, business intelligence, and exploratory data analysis.

Note: Currently, the healthcare-specific text analysis features described in this guide are only available for English texts.

In SAP Connected Health, text analysis comprises the following processes that are specifically developed for the medical domain:

- **Entity Extraction**: The process of identifying both general and clinical named entities, for example, people, dates, diseases, and procedures. This process eliminates the ‘noise’ in medical data by highlighting relevant information. Extracting these entities transforms unstructured medical texts into structured and usable data.

- **Fact Extraction**: A higher-level semantic process that links entities as facts in the medical domain. For example, determining whether a medical procedure is possible or factual and found in a positive or negative context, or discovering the relationship between a patient and specific medication.

Text analysis in SAP Connected Health relies on a set of files that include dictionaries and rules for processing the language of medical records. The text analysis tool extracts important information contained in unstructured texts and converts it into structured data that users can analyze and act upon.

The ultimate goal of text analysis in healthcare is to transform this unstructured textual data from multiple sources into a harmonized clinical data warehouse. This process includes identifying, classifying, and contextualizing medical information, such as the names of medication, diseases, symptoms, and procedures.

The text analysis module for healthcare and life sciences provides the following functionality:

- Comprehensive dictionary-based extraction of clinical entities using third-party ontologies (NCI metathesaurus, RxNorm, ICDO-3, and NCBI)
- Extraction of section headings typically found in electronic medical records
- Entity assertion extraction (modality, polarity, reference)
- Entity relation extraction
- Text analysis post-processing to tailor output to target application requirements

Text Analysis output can be tailored to a partner’s needs through simple configuration. SAP Connected Health provides the following configurations for text analysis:

- EXTRATION_CORE_MEDICAL
- EXTRATION_CORE_PATHOLOGY

For more details about general-purpose text analysis capabilities, such as domain-agnostic entity extraction (Person, Date, Locality, Organization, and so on), see the SAP HANA Text Analysis Language Reference Guide.
3.1 Overview of Clinical Entity Extraction

Clinical entity extraction is the process of discovering and presenting specific named entities that appear in an unstructured text.

Clinical entities denote the names of medications, procedures, diseases, anatomical sites, and so on, that can be extracted from text. An entity is defined as a pairing of a word (or sequence of words), its semantic type, and a standard form or code. For example, asthma/med_DIS/C28397 is an entity in which asthma is the extracted word, med_DIS is the semantic type, and C28397 the unique CUI code from the NCI metathesaurus.

Pathology entities denote the names of cancers (histology) and biomarkers used to diagnoses cancers. For example, HER2/med_GENE/2064 is an entity in which HER2 is the extracted word, med_GENE is the semantic type, and 2064 is the unique code from the NCBI ontology.

The language modules included with the SAP Connected Health contain system dictionaries that are based on the ontologies contained in the NCI metathesaurus, RxNorm, ICDO-3, or NCBI. The extraction process uses this list of entities to extract clinical entities from the text and to assign a unique code. The system classifies each extracted entity by entity type and assigns a unique code to each entity before presenting this metadata in a standardized format.

3.2 Overview of Clinical Fact Extraction

Clinical fact extraction is the process of discovering and presenting facts or relationships that occur in unstructured text, for example, the relationship between a patient and specific medication.

Clinical facts are extracted using rules consisting of patterns that define the expressions used to represent assertions and relations relevant for the healthcare domain.

Assertions about entities like **Factual**, **Proposed**, or **Possible** are extracted as clinical facts. For example, the sentence “The patient will undergo a colonoscopy to rule out further bleeding.” contains two clinical entities: colonoscopy/med_PROC_D and bleeding/med_SYMP, whose modality is Proposed and Possible, respectively.
4 EXTRACTION_CORE_MEDICAL

This configuration is used to extract medical entities and facts from clinical patient narratives. It is not specific to any particular medical sub-domain. When this configuration is applied, the extraction processes focus on entities such as medications, diseases, symptoms, procedures, anatomical sites, and laboratory tests, as well as the context in which these entities are found in terms of modality and polarity.

The following table lists the clinical entity types, their possible assertion values, and their relation values that this text analysis configuration can extract:

Table 1:

<table>
<thead>
<tr>
<th>Entity Types</th>
<th>Entity Assertions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease</td>
<td>Modality</td>
</tr>
<tr>
<td>Symptom</td>
<td>Factual</td>
</tr>
<tr>
<td>Anatomical site</td>
<td>Possible</td>
</tr>
<tr>
<td>Laboratory test</td>
<td>Conditional</td>
</tr>
<tr>
<td>Diagnostic procedure</td>
<td>Hypothetical</td>
</tr>
<tr>
<td>Therapeutic procedure</td>
<td>Proposed</td>
</tr>
<tr>
<td>Medication</td>
<td>Reference</td>
</tr>
<tr>
<td></td>
<td>Associated with patient</td>
</tr>
<tr>
<td></td>
<td>Not associated with patient</td>
</tr>
<tr>
<td></td>
<td>Describes patient’s family</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient allergic to medication</td>
</tr>
<tr>
<td>Patient not tolerating medication</td>
</tr>
</tbody>
</table>

Related Information

Clinical Entity Extraction [page 8]
Medical Section Headings Extraction [page 13]
Clinical Entity Assertions [page 18]
Clinical Entity Relations [page 21]
4.1 Clinical Entity Extraction

Clinical entity extraction is the process of discovering and presenting specific named entities that appear in an unstructured text. Extracting clinical entities from unstructured medical text tells us what the document is about, such as the symptoms, lab tests, and procedures described.

The language modules included with SAP Connected Health contain system dictionaries that are based on the ontologies contained in the NCI metathesaurus, RxNorm, ICDO-3, or NCBI. The extraction process uses this list of entities to extract clinical entities from the text and to assign a unique code. The system classifies each extracted entity by entity type and assigns a unique code to each entity before presenting this metadata in a standardized format.

The extraction process involves the following general steps:

- Processing and analyzing the text (tokenizing, stemming, part-of-speech tagging)
- Finding entities of interest
- Assigning entities to the appropriate semantic type
- Associating entities with a normalized code
- Presenting the metadata in a standard format

Clinical entity extraction is based on the following ontologies:

- NCI Metathesaurus for names and codes for diseases, symptoms, procedures, lab tests, and anatomical sites
- RxNorm for names and codes for medications
- ICDO-3 for names and codes for histologies
- NCBI for names and codes for genes

Each entity is defined as a grouping of a name, its type, and its normalized form. For example:

Table 2:

<table>
<thead>
<tr>
<th>Literal Form</th>
<th>Entity Type</th>
<th>Normalized Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tylenol PM</td>
<td>med_MED/RxNorm</td>
<td>220581@RxNorm</td>
</tr>
<tr>
<td>Chest X-Ray</td>
<td>med_PROC_D/NCIm</td>
<td>C38103@NCIm</td>
</tr>
<tr>
<td>Diabetes</td>
<td>med_DIS/NCIm</td>
<td>C0011847@NCIm</td>
</tr>
</tbody>
</table>

The normalized value is derived from RxNorm unique code for medications and from the NCI Metathesaurus CUI (Concept Unique Identifier) code for diseases, symptoms, laboratory tests, diagnostic procedures, therapeutic procedures and anatomical sites.

The name of the data source is appended to the entity type name as well as the normalized code used in that data source. For example, Simvastatin is extracted as entity type med_MED/RxNorm and normalized to 36567@RxNorm, whereas Pancreatitis is extracted as med_DIS/NCIm and normalized to C0030305@NCIm.
Related Information

med_MED/RxNorm [page 9]
med_DIS/NCIm [page 9]
med_SYMP/NCIm [page 10]
med_LAB_TEST/NCIm [page 11]
med_PROC_D/NCIm [page 11]
med_PROC_T/NCIm [page 12]
med_ANAT/NCIm [page 13]

4.1.1 med_MED/RxNorm

Corresponds to names of medications based on the RxNorm ontology.

The names of medication usually include the following information:

- Brand names (Lasix, Aspirin)
- Generics
- Ingredients (furosemide, vitamin E)
- Collective name for the group of medications (diuretics, vitamins)

Examples

- Coumadin
- Caltrate plus D
- Simvastatin
- Protonix
- ECASA
- Acetaminophen

4.1.2 med_DIS/NCIm

Corresponds to the names of diseases, disorders, or syndromes based on the NCI Metathesaurus ontologies.

Examples

- Stage II colon cancer
- Raynaud's syndrome
- Large cell lymphoma
- Influenza
- Anemia
- Subdural hematomas
- Malaria
- Pancreatic carcinoma
- Infected arthroplasty
- Cerebral palsy
- Osteonecrosis of right mandible
- Peripheral neuroathy
- Asthma
- Gastric mass
- Wart

4.1.3 med_SYMP/NCIm

Corresponds to names of signs or symptoms based on the NCI Metathesaurus ontologies.
A symptom is any subjective evidence of disease, while a sign is any objective evidence of disease.

Examples

- Abdominal pain
- Bad taste in the mouth
- Skin rash
- Jaw pain
- Nausea
- Weight loss
- Blurred vision
- Fatigue
- Fever
- Hot flashes
- Tingling feeling
- Vomiting
- Runny nose
- Skin redder than usual
- Insect allergy
4.1.4 med_LAB_TEST/NCIm

Corresponds to names of laboratory tests based on the NCI Metathesaurus ontologies.

Laboratory tests usually check samples of body fluids or tissues. They help in diagnosing medical conditions, planning treatments, and monitoring diseases.

Examples

- Blood culture
- Blood cholesterol
- Clotting factors
- Lead screening
- Hepatitis A
- Fasting insulin
- CBC (complete blood count)
- Kidney stone risk panel
- Lyme disease tests
- Pap smear
- RBC count (red blood cell count)
- Stool culture
- Urine analysis
- White count (white blood cell count)
- Alkaline phosphatase

4.1.5 med_PROC_D/NCIm

Corresponds to names of diagnostic procedures (other than laboratory tests) based on the NCI Metathesaurus ontologies.

Examples

- Cardiac stress test
- Endoscopy
- Magnetic resonance imaging
- Electrocardiography
- Esophageal motility study
- Positron emission tomography
- Electroencephalography
- Evoked potential
• Radiology
• Electroencephalography
• Magnetoencephalography
• Scintillography
• Electromyography
• Medical imaging (angiography)
• sPECT
• Electroneuronography
• Chest photofluorography
• Ultrasonography
• Electronystagmography
• Computed tomography
• Electrooculography
• Echocardiography
• Virtual colonoscopy
• Electroretinography
• Electrical impedance tomography
• Neuroimaging
• Endoluminal capsule monitoring
• Fluoroscopy
• Posturography

4.1.6 med_PROC_T/NCIm

Corresponds to names of therapeutic procedures based on the NCI Metathesaurus ontologies.

The aim of therapeutic procedures is to treat, cure, or restore function or structure in patients.

Examples

• Hysterectomy
• Right total wrist implant
• Liver transplant
• Angioplasty
• Intravenous hydration
• Empiric antibiotics
• Tetanus immunization
• Excision of palmar fascia
• Intravenous block anesthesia
• Appendectomy
• Mastectomy
• Radiation
4.1.7 med_ANAT/NCIm

Corresponds to names of anatomical sites based on the NCI Metathesaurus ontologies.

Only anatomical references not already included in other entity types are extracted. For example, the system does not extract the word knee in the phrase knee replacement surgery as type med_ANAT/NCIm. Instead, the whole phrase is extracted as type med_PROC_T/NCIm.

Examples

- Foot
- Left big toe
- Right foot sole
- Left thigh
- Vena cava
- Aorta
- Hair
- Tooth
- Skin

4.2 Medical Section Headings Extraction

Clinical entity extraction identifies and classifies the section headings that usually form the structure of medical reports, for example, History of Present Illness and Family History. As the same section heading can have multiple variants, each section heading entity has also been normalized. The variants also consider commonly misspelled words that can appear in these headings and normalize them accordingly.

The following table lists the entity types, normalized values, and definitions for section headings frequently found in electronic medical records, and also gives examples for each category:

<table>
<thead>
<tr>
<th>Entity Type</th>
<th>Definition</th>
<th>Normalized Form</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>med_SECTION_METADATA</td>
<td>Information about the docu-</td>
<td>METADATA</td>
<td>REPORT STATUS:</td>
</tr>
<tr>
<td></td>
<td>ment itself</td>
<td>NOTE TYPE</td>
<td>RECORD #</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DICTATED BY:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>COMPLETED BY:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>eScription document:</td>
</tr>
<tr>
<td>Entity Type</td>
<td>Definition</td>
<td>Normalized Form</td>
<td>Examples</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------</td>
<td>-------------------------------------------</td>
</tr>
<tr>
<td>med_SECTION_ADMISSION</td>
<td>Information about the patient's admission to a facility</td>
<td>ADMISSION</td>
<td>ADMISSION INTO THE HOSPITAL:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ADMIT:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ADMITTANCE:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ADMITTED:</td>
</tr>
<tr>
<td>med_SECTION_ALLERGIES</td>
<td>Information about allergies, adverse reactions, or drug interactions</td>
<td>ALLERGIES</td>
<td>ALLERGIC REACTION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DRUG ALLERGIES:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SERIOUS INTERACTION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ALL:</td>
</tr>
<tr>
<td>med_SECTION_ASSESSMENT</td>
<td>Information about the status of the patient or the impression the physician has of the patient</td>
<td>ASSESSMENT</td>
<td>DISPOSITION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>OVERALL IMPRESSION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PREOPERATIVE STATUS:</td>
</tr>
<tr>
<td>med_SECTIONCONSULTATION</td>
<td>Usually contains the name of the attending or referring physician</td>
<td>CONSULTATION</td>
<td>CONSULT:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATTENDING PHYSICIAN:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CONSULTANT ON CASE:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PHYSICIAN:</td>
</tr>
<tr>
<td>med_SECTION_DATE</td>
<td>Any section heading about dates or time</td>
<td>CONSULT DATE</td>
<td>DATE OF CONSULT:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADMISSION DATE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DISCHARGE DATE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SURGERY DATE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DATE OF DEATH</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DATE OF BIRTH</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>TEST DATE/TIME</td>
<td></td>
</tr>
<tr>
<td>med_SECTION_DIAGNOSIS</td>
<td>Information about the patient's diagnoses or medical problems</td>
<td>DIAGNOSIS</td>
<td>ASSOCIATED DIAGNOSIS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADMISSION DIAGNOSIS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DISCHARGE DIAGNOSIS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PREOPERATIVE DIAGNOSIS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>POSTOPERATIVE DIAGNOSIS</td>
<td></td>
</tr>
<tr>
<td>Entity Type</td>
<td>Definition</td>
<td>Normalized Form</td>
<td>Examples</td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------------------------------------------------------------------</td>
<td>------------------------------------------------------</td>
<td>---------------------------------------------------</td>
</tr>
<tr>
<td>med_SECTION_DIAG_DATA</td>
<td>Information about testing procedures, such as laboratory exams or imaging procedures at the time of admission, discharge, or unspecified</td>
<td>TESTS/LABS, TEST DATA ON ADMISSION, TEST DATA AT DISCHARGE</td>
<td>DIAGNOSTIC STUDIES: IMPORTANT FINDINGS: LABS/STUDIES: XRAY: DISCHARGE RESULTS: EKG:</td>
</tr>
<tr>
<td>med_SECTION_DISCHARGE</td>
<td>Information about the patient’s discharge from a facility</td>
<td>DISCHARGE, DEATH DETAILS</td>
<td>AT DISCHARGE: DISMISSAL: CONDITION AT DISCHARGE: CONDITION ON TRANSFER: CAUSE OF DEATH:</td>
</tr>
<tr>
<td>med_SECTION_FAMILY_HISTORY</td>
<td>Information about medical problems of the patient’s relatives</td>
<td>FAMILY HISTORY</td>
<td>FAM HISTORY: FH: FAMILY MEMBER DISEASE:</td>
</tr>
<tr>
<td>med_SECTION_FOLLOWUP</td>
<td>Instructions for patient to follow after being discharged from hospital or lists medical appointments to schedule</td>
<td>FOLLOWUP</td>
<td>ACTIVITY AT DISCHARGE: PLAN: DIET AT DISCHARGE: FOLLOWUP APPOINTMENTS: RECOMMENDATIONS: STATUS AT DISCHARGE:</td>
</tr>
<tr>
<td>med_SECTION_HISTORY</td>
<td>Information about the patient’s medical or surgical history or explaining the reason for the patient’s visit or hospitalization</td>
<td>MEDICAL HISTORY, SURGICAL HISTORY, CARDIO HISTORY, GYN HISTORY, REASON FOR SURGERY, REASON FOR VISIT</td>
<td>HISTORY OF PRESENT ILLNESS: HPI: PAST SURGICAL HISTORY: CHIEF COMPLAINT: REASON FOR ADMISSION: PRINCIPAL COMPLAINTS: PRIOR CARDIAC HISTORY: REASONS FOR PROCEDURE: INDICATIONS:</td>
</tr>
<tr>
<td>Entity Type</td>
<td>Definition</td>
<td>Normalized Form</td>
<td>Examples</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>--------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>med_SECTION_HOSPITAL_COURSE</td>
<td>Information about the patient's hospital stay</td>
<td>HOSPITAL COURSE</td>
<td>COURSE OF HOSPITALIZATION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HOSPITAL COURSE BY PROBLEM:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>POSTOPERATIVE COURSE:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HOSPITAL COURSE BY SYSTEM:</td>
</tr>
<tr>
<td>med_SECTION_MAIL_HEADER</td>
<td>Different types of specific headers commonly used to indicate sections within an e-mail.</td>
<td>MAIL</td>
<td>TO:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FROM:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CC:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>RE:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ATTN:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SUBJECT:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SUBJ:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>EMAIL:</td>
</tr>
<tr>
<td>med_SECTION_MEDS</td>
<td>Lists of the medications the patient is taking at home, on admission to, or discharge from hospital</td>
<td>MEDS</td>
<td>CURRENT MEDS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HOME MEDS</td>
<td>OUTPATIENT MEDS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ADMISSION MEDS</td>
<td>AT-HOME MEDS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DISCHARGE MEDS</td>
<td>MECDS AT ADMISSION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ANESTHESIA</td>
<td>MEDICATIONS AT TRANSFER:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MEDS INSTRUCTIONS:</td>
<td>DISCHARGE MEDICATIONS:</td>
</tr>
<tr>
<td>med_SECTION_PATIENT</td>
<td>General personal information about the patient</td>
<td>PATIENT</td>
<td>PATIENT NAME:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PATIENT GENDER</td>
<td>PATIENT ADDRESS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SEX:</td>
</tr>
<tr>
<td>Entity Type</td>
<td>Definition</td>
<td>Normalized Form</td>
<td>Examples</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>----------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>med SECTION PHYS EXAM</td>
<td>Information about physical examinations or review of systems</td>
<td>PHYSICAL EXAM</td>
<td>EXAMINATION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHYS EXAM ON ADMISSION</td>
<td>PHYSICAL EXAM AT ADMISSION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PHYS EXAM AT DISCHARGE</td>
<td>REVIEW OF SYSTEMS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>VITAL SIGNS</td>
<td>EXAM AT THE TIME OF DISCHARGE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HEMATOLOGICAL</td>
<td>VITALS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ABDOMEN</td>
<td>GI:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ENDOCRINE</td>
<td>GEN DESCRIPTION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EXTREMITIES</td>
<td>PELVIS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GENERAL APPEARANCE</td>
<td>CARDIAC:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GENITOURINARY</td>
<td>EARS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GYNECOLOGIC</td>
<td>NECK:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RECTAL</td>
<td>PULMONARY:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HEART</td>
<td>DERMATOLOGIC:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HEENT</td>
<td>PHYSICAL EXAMINATION ON ADMISSION TO THE INTENSIVE CARE UNIT:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NECK</td>
<td>PHYSICAL EXAMINATION ON ARRIVAL:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CHEST</td>
<td>PHYSICAL EXAM ON TRANSFER TO GENERAL MEDICAL FLOOR:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BACK</td>
<td>LEFT HUMERUS AND LEFT ELBOW:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LUNGS</td>
<td>ROS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LYMP NODES</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NEUROLOGIC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PSYCHIATRIC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RENAL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MUSCULOSKELETAL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SKIN</td>
<td></td>
</tr>
<tr>
<td>Entity Type</td>
<td>Definition</td>
<td>Normalized Form</td>
<td>Examples</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>med_SECTION_PROBLEM</td>
<td>Headings commonly used to indicate the start of sections that describe specific problems.</td>
<td>PROBLEM SPECIFICATION</td>
<td>ACIDOSIS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ACUTE PAIN:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ASYMPTOMATIC HYPOTENSION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BRADYCARDIA:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CARDIOGENIC SHOCK:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DIABETES</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HYPOGLYCEMIA:</td>
</tr>
<tr>
<td>med_SECTION_PROCEDURE</td>
<td>Information about surgical or therapeutic procedures the patient received</td>
<td>PROCEDURE</td>
<td>MEDICAL INTERVENTION:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CARDIAC PROCEDURE</td>
<td>ASSOCIATED SURGERIES:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROCEDURE DETAILS</td>
<td>INVASIVE PROCEDURE:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>COMPLICATIONS</td>
<td>OPERATION PERFORMED:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BLOOD LOSS/TRANSFUSION</td>
<td>SURGERIES DONE:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CARDIAC TRANSPLANT:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BLOOD LOSS:</td>
</tr>
<tr>
<td>med_SECTION_SOCIAL_HISTORY</td>
<td>Information about the patient’s smoking and drinking habits and psychosocial history</td>
<td>SOCIAL HISTORY</td>
<td>ALCOHOL ABUSE:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HABITS:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PSYCHOSOCIAL HISTORY:</td>
</tr>
</tbody>
</table>

### 4.3 Clinical Entity Assertions

Assertion extraction allows downstream applications to determine whether an entity should be included in analytical processes.

Assertion extraction provides the following contextual information about an entity:

- **Modality** Indicates whether the entity is factual, possible, conditional, hypothetical, or proposed.
- **Polarity** Indicates whether the entity is negated in the context in which it is found.
- **Reference** Indicates whether the entity is associated with the patient or a member of the patient’s family.

### Related Information

- Modality Assertion [page 19]
4.3.1 Modality Assertion

There are five types of modality assertion.

Table 4: Types of Modality Assertion

<table>
<thead>
<tr>
<th>Modality</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>ModFactual</td>
<td>Any clinical entity found in a context not expressing possibility, uncertainty, or conditionality.</td>
<td>• The patient was diagnosed with <em>asthma</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <em>X-ray</em> was done yesterday.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• History of <em>pneumonia</em> 2 years ago.</td>
</tr>
<tr>
<td>ModPossible</td>
<td>Any clinical entity found in a context expressing possibility or uncertainty.</td>
<td>• We suspect this is not <em>pneumonia</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Small chance of <em>pneumonia</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <em>Pneumonia</em> unlikely.</td>
</tr>
<tr>
<td>ModConditional</td>
<td>Any clinical entity found in a context where the patient experiences the entity (mostly symptoms or diseases) only in certain conditions.</td>
<td>• Patient has had increasing dyspnea on exertion.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Penicillin causes a <em>rash</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Patient reports shortness of <em>breath</em> upon climbing stairs.</td>
</tr>
<tr>
<td>ModHypothetical</td>
<td>Any clinical entity found in a hypothetical entity (including expressions such as &quot;PRN&quot;)</td>
<td>• If there is evidence of <em>metastatic disease</em>, then potential medical therapy would include chemotherapy radiation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <em>TYLENOL</em> 650 MG PO Q6H PRN <em>pain</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Please recheck fingerstick in 1 hour if administering <em>NovoLog</em> at bedtime.</td>
</tr>
<tr>
<td>ModProposed</td>
<td>Any clinical entity that is proposed or recommended.</td>
<td>• Follow up echocardiogram in 2 weeks.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Dr. Smith recommended <em>gastroscopey</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Patient was advised about angio-<em>plasty</em>.</td>
</tr>
</tbody>
</table>
4.3.2 Polarity Assertion

There are two types of polarity assertion.

Table 5: Types of Polarity Assertion

<table>
<thead>
<tr>
<th>Polarity</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoPositive</td>
<td>Any clinical entity that is found in a non-negative or affirmative context.</td>
<td>• The patient was diagnosed with <em>asthma</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <em>X-ray</em> was done yesterday.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• History of <em>pneumonia</em> 2 years ago.</td>
</tr>
<tr>
<td>PoNegative</td>
<td>Any clinical entity that is found in a negative context, for example, a disease a patient does not have or a medication the patient has stopped.</td>
<td>• I instructed her to withhold <em>Coumadin</em> for two days.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Patient denies <em>night sweats</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <em>EKG</em> was not done.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Patient declined <em>angioplasty</em>.</td>
</tr>
</tbody>
</table>

4.3.3 Reference Assertion

There are three types of reference assertion.

Table 6: Types of Reference Assertion

<table>
<thead>
<tr>
<th>Reference</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>med_RefAssociated</td>
<td>Any clinical entity associated with a patient.</td>
<td>• The patient was diagnosed with <em>asthma</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• <em>X-ray</em> was done yesterday.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Patient has a history of <em>pneumonia</em>.</td>
</tr>
<tr>
<td>med_RefNotAssociated</td>
<td>Any clinical entity not associated with the patient or patient’s family.</td>
<td>• He was given a brochure about the risks of <em>anti-hypertensives</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Patient was given information about <em>anemia</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Post-surgery risk of <em>abdominal pain</em> was explained to the patient.</td>
</tr>
<tr>
<td>med_RefFamily</td>
<td>Any clinical entity associated with the patient’s family.</td>
<td>• The patient’s brother has been on <em>anti-hypertensives</em> for 10 years.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• The patient’s father has also complained of <em>abdominal pain</em>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mother had <em>liver transplant</em>.</td>
</tr>
</tbody>
</table>
4.4 Clinical Entity Relations

Relation extraction provides allergy or intolerance information about an entity. For example, information about whether the entity causes, or is currently causing, an allergic or intolerance reaction for the patient or a member of the patient’s family.

Relation extraction for clinical reports can link the following types of information:

- Allergy information about a medication or a food item
- Intolerance information about a medication or a food item
- Detailed medication information which includes the name of a medication along with its strength, doseform, dosage, frequency, route, and/or reason for taking the medication

The links between these types of information are extracted as the following types of relations:

Table 7: Types of Relation

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>med_RelAllergic</td>
<td>There is an allergy relationship.</td>
<td>The patient is allergic to Tylenol.</td>
</tr>
<tr>
<td>med_RelIntolerant</td>
<td>There is an intolerance relationship.</td>
<td>Patient did not tolerate penicillin.</td>
</tr>
</tbody>
</table>
| med_Medication    | Detailed medication information: name of medication along with its strength, dose form, dosage, frequency, route, and/or reason for taking the medication | MEDS:  
1. Diltiazem sustained release 120 mg p.o. daily.  
2. Caltrate 600 plus D one tablet p.o. daily. |
This configuration is used to extract cancer-related entities and facts from pathology reports. It is specific to the oncology and pathology medical domains. When this configuration is applied, the extraction processes focus on two types of entities:

- Cancer-related entities, such as histology, grade, or stage and the context in which these entities are found in terms of modality and polarity
- Biomarker-related entities, such as tests used, results, or genetic alteration types and locations, and the context in which these entities are found in terms of modality and polarity

The following table lists the clinical entity types, their possible assertion values, and their relation values that this text analysis configuration can extract:

<table>
<thead>
<tr>
<th>Entity Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene</td>
</tr>
<tr>
<td>Histology</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity Assertions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modality</td>
</tr>
<tr>
<td>○ Factual</td>
</tr>
<tr>
<td>○ Possible</td>
</tr>
<tr>
<td>○ Conditional</td>
</tr>
<tr>
<td>○ Hypothetical</td>
</tr>
<tr>
<td>○ Proposed</td>
</tr>
<tr>
<td>Polarity</td>
</tr>
<tr>
<td>○ Positive</td>
</tr>
<tr>
<td>○ Negative</td>
</tr>
<tr>
<td>Reference</td>
</tr>
<tr>
<td>○ Associated with patient</td>
</tr>
<tr>
<td>○ Not associated with patient</td>
</tr>
<tr>
<td>○ Describes patient’s family</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Entity Relations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer name: histology, behavior, anatomy</td>
</tr>
<tr>
<td>Cancer grade: type and level</td>
</tr>
<tr>
<td>Cancer stage: type and level</td>
</tr>
<tr>
<td>Cancer size: tumor size</td>
</tr>
<tr>
<td>Biomarker name: test result, test name, gene/protein alteration, location of alteration</td>
</tr>
</tbody>
</table>

Related Information

Pathology Entity Extraction [page 23]
Medical Section Headings Extraction for Pathology [page 24]
5.1 Pathology Entity Extraction

Extracting pathology entities from unstructured medical text tells us what the text is about, such as the cancer names, grades, stages, and biomarkers used to diagnose the tumor.

The extraction process involves the following general steps:

- Processing and analyzing the text (tokenizing, stemming, part-of-speech tagging)
- Finding entities of interest
- Assigning entities to the appropriate semantic type
- Associating entities with a normalized code
- Presenting the metadata in a standard format

Pathology entity extraction is based on the ontologies from the ICDO-3 (histology) and the NCBI (gene).

Each entity is defined as a grouping of a name, its semantic type, and its normalized form. For example:

Table 9:

<table>
<thead>
<tr>
<th>Literal Form</th>
<th>Entity Type</th>
<th>Normalized Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>squamous intraepithelial neoplasia</td>
<td>med_HISTO/ICDO3</td>
<td>80772@ICDO3</td>
</tr>
<tr>
<td>HER2/neu</td>
<td>med_GENE/NCBI</td>
<td>2064@NCBI</td>
</tr>
</tbody>
</table>

The normalized value is derived from the ICDO-3 unique code for cancer names and from the NCBI code for genes. The name of the data source is appended to the entity type name as well as the normalized code used in that data source.

Related Information

med_HISTO/ICDO3 [page 23]
med_GENE/NCBI [page 24]

5.1.1 med_HISTO/ICDO3

Corresponds to names of cancers based on the ICDO-3 ontology.

The names of cancers usually include the following information:

- Type of tumor (carcinoma, sarcoma, myeloma, leukemia, lymphoma, mixed type)
• Behavior (benign, borderline, in situ, malignant)

Examples

• Infiltrating ductal carcinoma
• Neuroendocrine carcinoma
• Malignant neoplasm
• Spindle cell sarcoma

5.1.2 med_GENE/NCBI

Corresponds to the names of genes based on the NCBI ontology.

Examples

• HER2
• ALK
• PSA
• KRAS

5.2 Medical Section Headings Extraction for Pathology

For the text analysis configuration EXTRACTION_CORE_PATHOLOGY, medical section headings extraction works the same as it does for the EXTRACTION_CORE_MEDICAL configuration.

For more information, see Medical Section Headings Extraction [page 13].

5.3 Pathology Entity Assertions

Assertion extraction allows downstream applications to determine whether an entity should be included in analytical processes.

For the text analysis configuration EXTRACTION_CORE_PATHOLOGY, entity assertions extraction works the same as it does for the EXTRACTION_CORE_MEDICAL configuration, only that it is for pathology-related entities. For more information, see Clinical Entity Assertions [page 18].
Assertion extraction for pathology entities provides the following information about an entity:

- **Modality**
- **Polarity**
- **Reference**

**Modality**

Modality indicates whether the entity is factual, possible, conditional, hypothetical, or proposed.

There are five types of modality assertion:

- ModFactual
- ModPossible
- ModConditional
- ModHypothetical
- ModProposed

**Polarity**

Polarity indicates whether the entity is negated in the context in which it is found.

There are two types of polarity assertion:

- PolPositive
- PolNegative

**Reference**

Indicates whether the entity is associated with the patient or a member of the patient’s family.

There are three types of reference assertion:

- med_RefAssociated
- med_RefNotAssociated
- med_RefFamily

### 5.4 Pathology Entity Relations

Relation extraction for pathology reports can link the following types of information:

- Cancer name with its behavior and anatomical site
• Cancer grade and its value
• Cancer stage and its value
• Cancer tumor size
• Biomarker, result, test used, and gene alteration

The links between these types of information are extracted as the following types of relations:

Table 10:

<table>
<thead>
<tr>
<th>Relation</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>med_CancerHistology</td>
<td>Extracts a span of text referring to histology.</td>
<td>• Acute myeloid leukemia M6 type&lt;br&gt;• Squamous intraepithelial neoplasia&lt;br&gt;• Invasive lung carcinoma</td>
</tr>
<tr>
<td></td>
<td>It may include behavior and anatomical site information as well.</td>
<td></td>
</tr>
<tr>
<td>med_CancerGrade</td>
<td>Extracts a span of text that refers to a cancer grade and its value. The value can be expressed in words (well differentiated, undifferentiated) or in numbers (grade 1, Gleason 6, G3).</td>
<td>• Well-differentiated invasive lung carcinoma&lt;br&gt;• GLEASON SCORE 6 (3 + 3)</td>
</tr>
<tr>
<td>med_CancerStage</td>
<td>Extracts a span of text that refers to a cancer stage and its value. The cancer stage is usually expressed using the TNM model (tumor, node, and metastasis).</td>
<td>• PATHOLOGIC STAGING: pT1c, pN0(sn) (STAGE IA)&lt;br&gt;• Regional Lymph Nodes (pN): pN0</td>
</tr>
<tr>
<td>med_CancerSize</td>
<td>Extracts a span of text that refers to a cancer size.</td>
<td>• Invasive tumor size: 1.0 cm&lt;br&gt;• Size of invasive carcinoma: 1.7 x 1.0 x 1.0 cm</td>
</tr>
<tr>
<td>med_Biomarker</td>
<td>Extracts a span of text that refers to a biomarker and the information associated with the biomarker, such as:</td>
<td>• EGFR is positive for mutation in exon 19&lt;br&gt;• HER2/neu positive, score 3+&lt;br&gt;• Positive for fusion in ALK gene&lt;br&gt;• ERBB2 CISH equivocal&lt;br&gt;• Immunohistochemical stains are positive for CAM 5.2 and negative for CD45</td>
</tr>
<tr>
<td></td>
<td>• Test used in assessing biomarker (FISH, immunohistochemistry)&lt;br&gt;• Biomarker test result (positive, negative, borderline, etc.) and value/score of test if present (3+, 2, etc.)&lt;br&gt;• Type of alteration the test is measuring (fusion, amplification, mutation, translocation)&lt;br&gt;• Location of alteration if present (exon, codon)</td>
<td></td>
</tr>
</tbody>
</table>
Important Disclaimers and Legal Information

Coding Samples

Any software coding and/or code lines / strings (“Code”) included in this documentation are only examples and are not intended to be used in a productive system environment. The Code is only intended to better explain and visualize the syntax and phrasing rules of certain coding. SAP does not warrant the correctness and completeness of the Code given herein, and SAP shall not be liable for errors or damages caused by the usage of the Code, unless damages were caused by SAP intentionally or by SAP’s gross negligence.

Accessibility

The information contained in the SAP documentation represents SAP’s current view of accessibility criteria as of the date of publication; it is in no way intended to be a binding guideline on how to ensure accessibility of software products. SAP in particular disclaims any liability in relation to this document. This disclaimer, however, does not apply in cases of willful misconduct or gross negligence of SAP. Furthermore, this document does not result in any direct or indirect contractual obligations of SAP.

Gender-Neutral Language

As far as possible, SAP documentation is gender neutral. Depending on the context, the reader is addressed directly with “you”, or a gender-neutral noun (such as “sales person” or “working days”) is used. If when referring to members of both sexes, however, the third person singular cannot be avoided or a gender-neutral noun does not exist, SAP reserves the right to use the masculine form of the noun and pronoun. This is to ensure that the documentation remains comprehensible.

Internet Hyperlinks

The SAP documentation may contain hyperlinks to the Internet. These hyperlinks are intended to serve as a hint about where to find related information. SAP does not warrant the availability and correctness of this related information or the ability of this information to serve a particular purpose. SAP shall not be liable for any damages caused by the use of related information unless damages have been caused by SAP’s gross negligence or willful misconduct. All links are categorized for transparency (see: http://help.sap.com/disclaimer).