Linguamatics NLP Precision Oncology: Transforming Clinical Documentation Into Cancer Insights

Challenge

Rich clinical data sets are essential in cancer to drive clinical care and research. Diagnosis, treatment selection, predictive models, analytics, data warehouses, tissue banks and cancer registries all rely on detailed cancer data. However, the collection of this data is often slow and manual because it is trapped in complex documentation such as pathology, imaging and genetic reports.

Solution

Linguamatics NLP for Healthcare automatically extracts key cancer insights from clinical documentation to improve hospital efficiency and patient outcomes. Unlike open source and clinical coding packages, Linguamatics NLP for Healthcare combines the usability and speed of a search engine with the full power of natural language processing (NLP) to extract a wide range of cancer characteristics such as tumor histology, topography, biomarkers and stage. The product is easily configurable to extract additional concepts.

Benefits

By applying Linguamatics NLP for Healthcare, hospitals and specialist cancer centers can:

• extract pathology data in real time to populate electronic health records (EHRs) with detailed, up-to-date cancer information for tumor boards and other clinical decision-making
• save time and effort on manual extraction of data from clinical documentation
• identify patients for specific clinical trials by matching eligibility criteria to cancer characteristic
• work with the growing community of Linguamatics customers in cancer research and care
• drive diagnostic, prognostic and treatment models from a wider range of pathology, imaging, laboratory and genetic data
• add value to existing investments in Clinical Trials Management Systems, tissue banks and research warehouses with additional, detailed clinical information.
KEY FEATURES

- Automatically assign ICD-O codes from pathology reports.
- Integrate extraction queries into EHRs and Extract, Transform and Load (ETL) pipelines via RESTful Web Services APIs.
- Automatically extract biomarker values for ER, PR, HER2, ALK and EGFR, and easily extend.
- Support cancer-case finding in pathology reports.
- Automatically extract tumor characteristics such as histology, grade, behavior, body site, excision margin, size, stage and TNM stage.
- Run pre-built queries interactively, and build and modify new algorithms.
- Mine medical literature for associations between genetic variants and rare diseases.
- Automatically monitor for high-risk patients, for example pulmonary nodules mentioned in imaging reports, and flag them for follow-up.

Figure 1: Screenshot Of Linguamatics NLP For Healthcare Extracting Topography, Morphology (Histology, Behavior, Grade) And Laterality From The Cancer Genome Atlas (TCGA) Pathology Reports

<table>
<thead>
<tr>
<th>Topography</th>
<th>Histology, benign to Neoplasm, malignant, uncertain whether primary or metastatic</th>
<th>Behavior</th>
<th>Grade</th>
<th>Laterality</th>
<th>ICD-O Code</th>
<th>#Docs</th>
<th>#Hits</th>
<th>Hit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ascending colon</td>
<td>Neoplasm, benign to Neoplasm, malignant, uncertain whether primary or metastatic</td>
<td>Malignant, primary site</td>
<td>C18.2,M8000/3</td>
<td>1</td>
<td>TCGA-A6-2983, 06B91c4d-9b74-4c4b-8849-c866680da0a1</td>
<td></td>
<td></td>
<td>1 CLINICAL NOTES PRE-OP DIAGNOSIS: Ascending colon cancer</td>
</tr>
<tr>
<td>Base of tongue</td>
<td>Squamous cell carcinoma in situ to Squamous cell carcinoma, metastatic</td>
<td>Malignant, primary site</td>
<td>Grade II</td>
<td>Left</td>
<td>C01.9,M8070/2L</td>
<td>1</td>
<td></td>
<td>1: A: Base of tongue, Left biopsy, Invasive squamous cell carcinoma moderately differentiated</td>
</tr>
<tr>
<td>Brain</td>
<td>Glialastoma</td>
<td>Malignant, primary site</td>
<td>Grade IV</td>
<td>C71.9,M9440/34</td>
<td>TCGA-26-1438-49167b-548b-4560-9140-543ae079d0f8</td>
<td></td>
<td></td>
<td>1: A: Brain tumor, biopsy, Glialastoma multiforme (WHO Grade IV astrocytoma) (see comment)</td>
</tr>
<tr>
<td>Breast</td>
<td>Intrudcted carcinoma, noninfiltrating to infiltrating duct carcinoma</td>
<td>Malignant, primary site</td>
<td>Grade IV</td>
<td>C50.9,M8500/3</td>
<td>TCGA-An-AAMC-12638FFC-482e-49b2-8c4b-87805c437</td>
<td></td>
<td></td>
<td>1 Diagnosis: Breast Cancer Histological description: Infiltrative duct carcinoma</td>
</tr>
<tr>
<td>Colon</td>
<td>Neoplasm, benign to Neoplasm, malignant, uncertain whether primary or metastatic</td>
<td>Malignant, primary site</td>
<td>C18.9,M8000/3</td>
<td>1</td>
<td>TCGA-04-6311-9b73836d-50f6-40f9-9b5d-9efc45f1318</td>
<td></td>
<td></td>
<td>1 PRE-OPERATIVE DIAGNOSIS: Colon cancer ADDENDUM TO ANALYSIS OF MISMATCH ...</td>
</tr>
<tr>
<td>Frontal lobe</td>
<td>Glialastoma</td>
<td>Malignant, primary site</td>
<td>Left</td>
<td>C71.1,M9440/3L</td>
<td>TCGA-06-0125-299b8ae-2df5-909f-5b2a-5b5a8a35fa1</td>
<td></td>
<td></td>
<td>1 BRAIN BIOPSY, LEFT FRONTAL LOBE, GLIALASTOMA MULTIFORME</td>
</tr>
<tr>
<td>Hepatic flexure of colon</td>
<td>Adenoma to Adenocarcinoma, metastatic</td>
<td>Malignant, primary site</td>
<td>Grade II</td>
<td>C18.3,M9170/32</td>
<td>TCGA-A6-2890-Sep27b7-6127-4c6d-9112-876170a10d</td>
<td></td>
<td></td>
<td>1 Colon, hepatic flexure, extended hemioclacomy: Adenocarcinoma 1 moderately-differentiated with invasion through the ...</td>
</tr>
</tbody>
</table>

CONTACT US

+44 (0)1223 651 910 (U.K.) | +1 617 674 3256 (U.S.)
nlp@iqvia.com
linguamatics.com