

INTRODUCTION

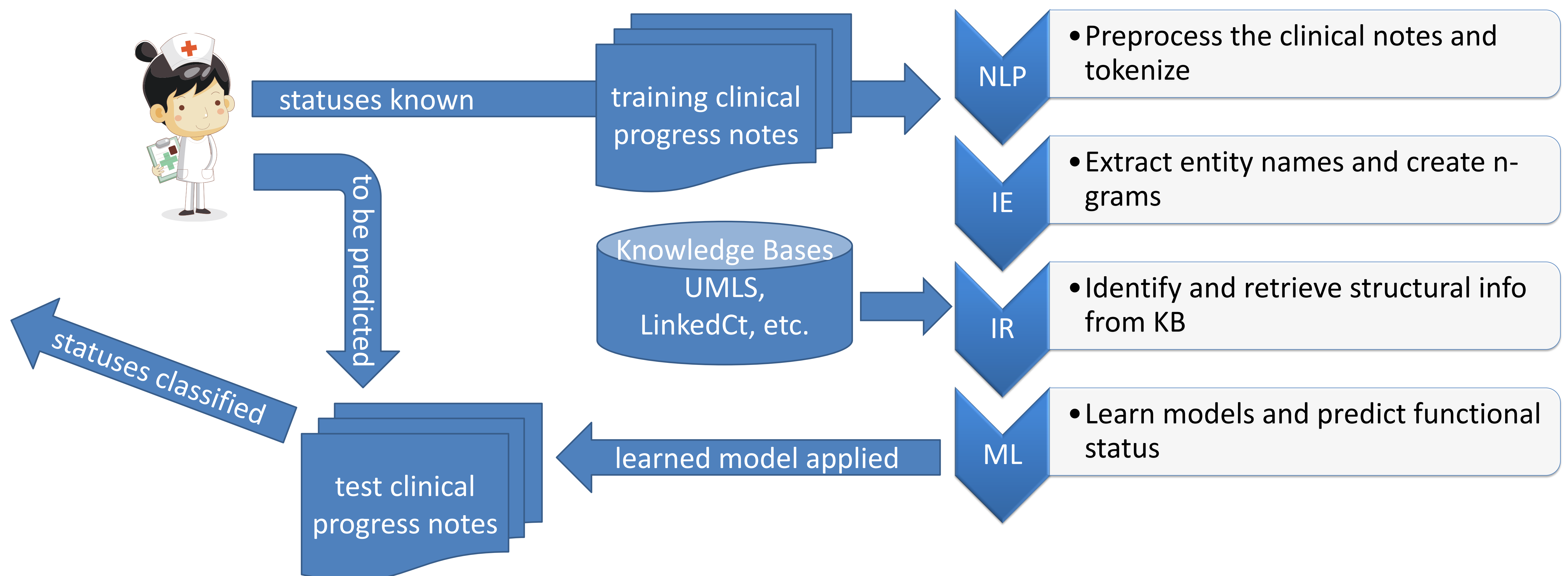
Medical records contain free texts like progress notes where healthcare professionals record details of a patient's clinical status. Progress notes document the care delivered and relevant clinical events to diagnosis and treatment for a patient, and they constitute the vast majority of medical records [1].

Number and size of biomedical data sources in computer readable format are increasing rapidly. For example while UMLS brings together over 100 vocabularies [2] (some of which are ICD-9, MeSH, OMIM and RxNorm), a new ontology LinkedCT is specialized only for clinical trials [3].

OBJECTIVE

Our goal is to predict functional status of patients by leveraging Information Retrieval (IR), Natural Language Processing (NLP), Information Extraction (IE), and Machine Learning (ML) on clinical free texts such as progress notes and biomedical domain ontologies such as UMLS.

Eating	✓
Transferring	✓
Grooming	✓
Toileting	✓
Bathing	✓
Walking	✓
Dressing	✓
Bowel Incontinence	✗
Bladder Control	✓



RELATED WORK

- MedLEE was an NLP system that extracts information from clinical narratives and presents this information in structured form using a controlled vocabulary. It has been successfully used to process radiology reports, discharge summaries, sign-out notes, etc. (Columbia U. commercialized it in Jan. 2012)[4]
- Since 2006 i2b2 organizes Workshop on Natural Language Processing Challenges for Clinical Records [5] which result in valuable academic publications.
- Brigham and Women's Hospital and Harvard Medical School developed Health Information Text Extraction (HiTex) tool for extracting specific findings from clinical narrative [6].

MOTIVATION

- Machine Learning in Healthcare research has been mainly focused on using structured data. However, unstructured free texts in healthcare also contain overwhelming information which can be utilized by text mining. It is also important to utilize databases and ontologies to retrieve information of related concepts to patients' functional status.
- Measuring activities of daily living (ADLs) is done repeatedly in chronic disease hospitals to monitor the improvement of patients and is also a requirement for nursing homes. Collection of functional status of patients are not only time consuming, but also accurate evaluations by clinicians might be difficult.
- Functional assessments are not available in some occasions like Veterans Administration Medical Foster Home (MFH) program.

METHOD

- StanfordNLP, UIMA, LingPipe are some of the text analysis tools (APIs) in Java which can be used in Text Classification, Chunk Recognition and Relation Extraction tasks.
- ML is applied to the entire preprocessed training dataset to learn models for predicting patients' functional status. Among applicable approaches that can be used are Naïve Bayes (NB) algorithm and K-Nearest Neighbor (KNN) classification methods.

REFERENCES

- [1] http://en.wikipedia.org/wiki/Progress_notes
- [2] http://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/source_vocabularies.html
- [3] <http://arxiv.org/ftp/arxiv/papers/0908/0908.0567.pdf>
- [4] <http://techventures.columbia.edu/news/columbia-grants-health-fidelity-exclusive-license-medlee-nlp>
- [5] <https://www.i2b2.org/NLP/DataSets/Main.php>
- [6] https://www.i2b2.org/software/projects/hitex/hitex_manual.html

