Logistics

- Grouping information for *ALL*
  - Name your group (of 6), specify each members (names & UFID's) by Wed. 5pm
  - Specify if to participate in NIST pre-pilot (smart city) in place of lab 4 and final project or NOT
  - Smaller groups of 2 to be formed for lab 3
- UF DSI & Python bootcamps
- UF Shands looking for part time on data visualization (experience, interest?)
- Lab 2 deadline (Thursday 5pm)
- Pop quiz coming up

Review

- Similarity metrics for approximate matching
  - Text string
  - Sets
  - Vector
- NLP in Biomedical domain with cTakes
  - Application: Biomedical data warehouse
  - NLP pipelines, tasks and methods
  - cTakes, UIMA and UMLS

Medication CEM template

- associatedCode
- Change_status
- Conditional
- Dosage
- Duration
- End_date
- Form
- Frequency
- Generic
- Negation_indicator
- Route
- Start_date
- Strength
- Subject
- Uncertainty_indicator

Procedure CEM template

- associatedCode
- Body_laterality
- Body_location
- Body_side
- Conditional
- Device
- End_date
- Generic
- Method
- Negation_indicator
- Relative_temporal_context
- Start_date
- Subject
- Uncertainty_indicator

Sign/Symptom CEM template

- Alleviating_factor
- associatedCode
- Body_laterality
- Body_location
- Body_side
- Conditional
- Course
- Duration
- End_time
- Exacerbating_factor
- Generic
- Negation_indicator
- Relative_temporal_context
- Severity
- Start_time
- Subject
- Uncertainty_indicator

Lab CEM template

- Abnormal_interpretation
- associatedCode
- Conditional
- Delta_flag
- Estimated_flag
- Generic
- Lab_value
- Negation_indicator
- Ordinal_interpretation
- Reference_range_narrative
- Subject
- Uncertainty_indicator

Disease/Disorder CEM template

- Alleviating_factor
- Associated_sign_or_symptom
- associatedCode
- Body_laterality
- Body_location
- Body_side
- Conditional
- Course
- Duration
- End_time
- Exacerbating_factor
- Generic
- Negation_indicator
- Relative_temporal_context
- Severity
- Start_time
- Subject
- Uncertainty_indicator

Anatomical Site CEM template

- associatedCode
- Body_laterality
- Body_site
- Conditional
- Generic
- Negation_indicator
- Subject
- Uncertainty_indicator

cTakes Example: Drug Object

- "Tamoxifen 20 mg po daily started on March 1, 2005."
  - Drug
    - Text: Tamoxifen
    - Associated code: C0351245
    - Strength: 20 mg
    - Start date: March 1, 2005
    - End date: null
    - Dosage: 1.0
    - Frequency: 1.0
    - Frequency unit: daily
    - Duration: null
    - Route: Enteral Oral
    - Form: null
    - Status: current
    - Change Status: no change
    - Certainty: null

cTakes Example: Disorder Object

- "No evidence of cholangiocarcinoma."
  - Disorder
    - Text: cholangiocarcinoma
    - Associated code: SNOMED 70179006
    - Certainty: 1
    - Context: current
    - Relatedness to patient: true
    - Status: negated

Applications in Biomedicine

- Translational science and clinical investigation
  - Patient cohort identification
  - Phenotype extraction
  - Linking patient's phenotype and genotype
  - eMERGE, PGRN, i2b2, SHARP
- Meaningful use of the EMR
- Comparative effectiveness
- Epidemiology
- Clinical practice
- …..

Biomedical NLP/KBP Research @ UF Data Science Research Group

- Knowledge extraction and outcome prediction of surgeries based on EHR notes
  - Phenotype Knowledge base (PheKB) is a collaborative environment to building and validating electronic algorithms to identify phenotype characteristics of patients within health data using i2b2 facts and NLP techniques.
- Knowledge exchange between doctors and probabilistic prediction models
- Biomedical Big Data Science Research
  - UF CISE three-course series on Data Science