SPARK NLP IN ACTION:

HOW SELECTDATA USES AI TO BETTER UNDERSTAND HOME HEALTH PATIENTS

Alberto Andreotti
Stacy Ashworth
Tawny Nichols
David Talby

John Snow LABS

SelectData™

CONTENTS

1.

The Opportunity

HOW HOME HEALTH WORKS

- Health care services that can be given in your home
- Examples of skilled home health services include:
  - Wound care for pressure sores or a surgical wound
  - Patient and caregiver education
  - Intravenous or nutrition therapy
  - Injections
  - Monitoring serious illness and unstable health status
  - Therapy

- **The Goal** = Become as self-sufficient as possible

HOW HOME HEALTH WORKS

The Players

- Home-bound Patients
- Providers
  - Agencies, Doctors, and Clinicians collaborate to provide care
- Payers
  - The Centers for Medicare and Medicaid Services (CMS) pays a predetermined base payment with an adjustment for health conditions
    - Outcome and Assessment Information Set determines payment
    - Payment for the 60-day Episode (consolidated billing)
  - Commercial Insurance pays by service or visit

- **Bottom line:** Medical coding determines the correct code for billing a claim **AND** articulating known health conditions for treatment

MEDICAL CODING SHORTAGE

3M Reported:

- Lack of qualified candidates
- Large scale retirement
- Widespread electronic health record adoption
- Baby boomers

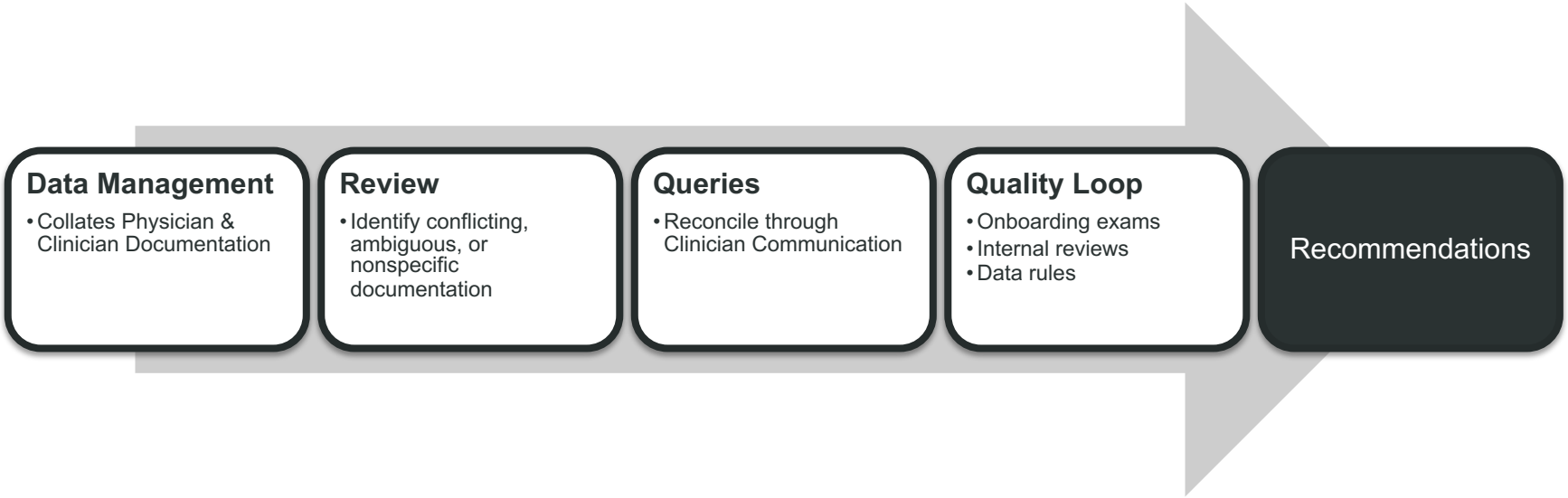
Coding backlogs:

- Loss of operating income
- Increase of denials or over/under payments
- Risk of audits or fines form CMS

HOME HEALTH CHALLENGES

- Many More Members
  - Today: 47 Million
  - 2030: 77 Million
- Skilled Worker Shortage
  - 1960s: 4.0 per member
  - Today: 3.4 per member
  - 2030: 2.3 per member
- Rising Costs from increased average life expectancy
  - 1960s: 68
  - Today: 78
  - 2030: 81

BUILDING THE NEXT GENERATION



HOME HEALTH CHALLENGES

Determining skill

How do we measure skill?
How do we control quality?

Too much information

- Where do we find relevant information?
- Is the information complete?

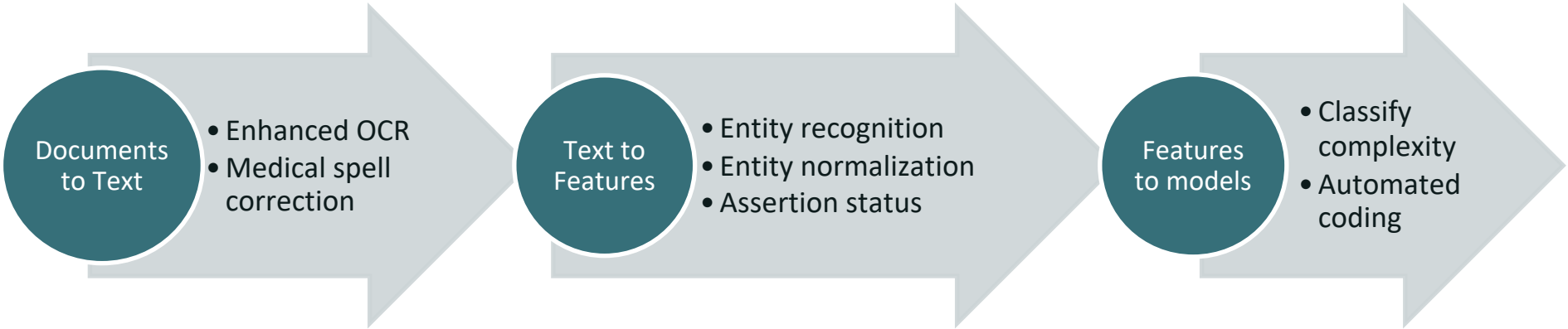
The path of least resistance

- Cognitive Dissonance

CHALLENGES

- Coding accuracy of 95%
  - Record classification
    - Over 14,000 ICD-10 Diagnosis Codes
    - Limitless referrals sources and formatting of similar data points
    - 60+ Nursing Specialties and/or observation perspectives (domain expertise)
  - Coder Skill classification
    - 21 Chapters with different types of injuries, factors, diseases, and symptoms
    - Knowledge on level of detail for 3, 4, 5, 6 or 7 characters (specificity)
    - Knowledge of chronic diseases and/ or coexisting conditions
  - Quality Assurance
    - Review % of overall records

SOLUTION BLUEPRINT



2.

From paper to digital text

NLP FOR APACHE SPARK

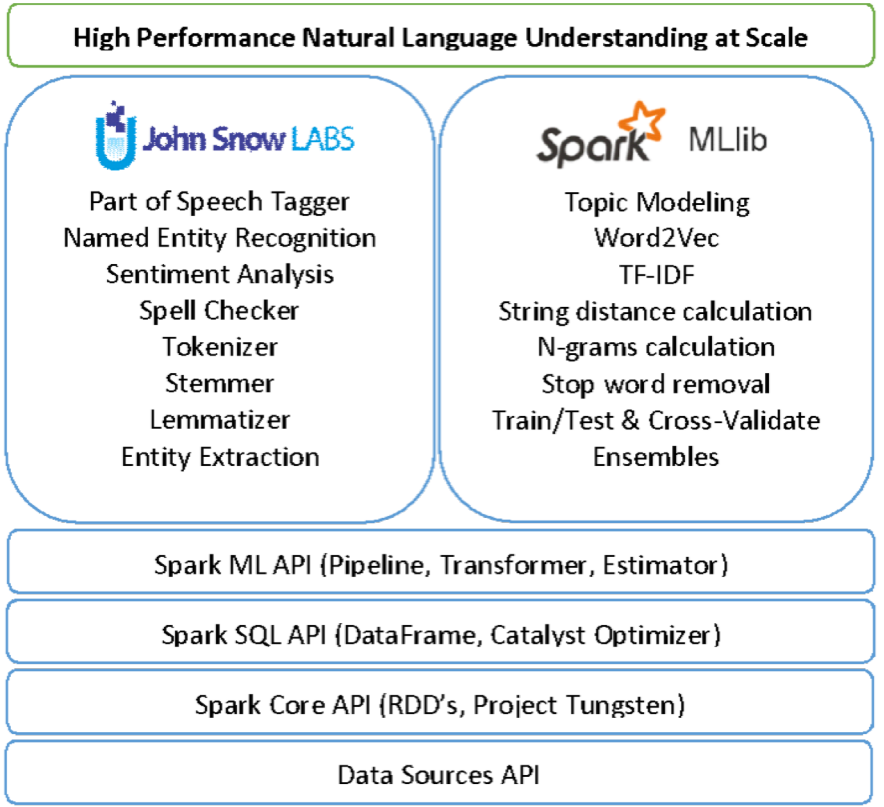
Design Goals

- State of the art Performance & Scale
- Frictionless Reuse
- Enterprise Grade

Built on the Spark ML API's

Apache 2.0 Licensed

Active development & support



NLP FOR APACHE SPARK: COMBINED NLP & ML PIPELINES

```
pipeline = pyspark.ml.Pipeline(stages=[
              document_assembler,
              tokenizer,
              stemmer,
              normalizer,
              stopword_remover,
              tf,
              idf,
              lda])

topic_model = pipeline.fit(df)
```

Spark NLP annotators

Spark ML featurizers

Spark ML LDA implementation

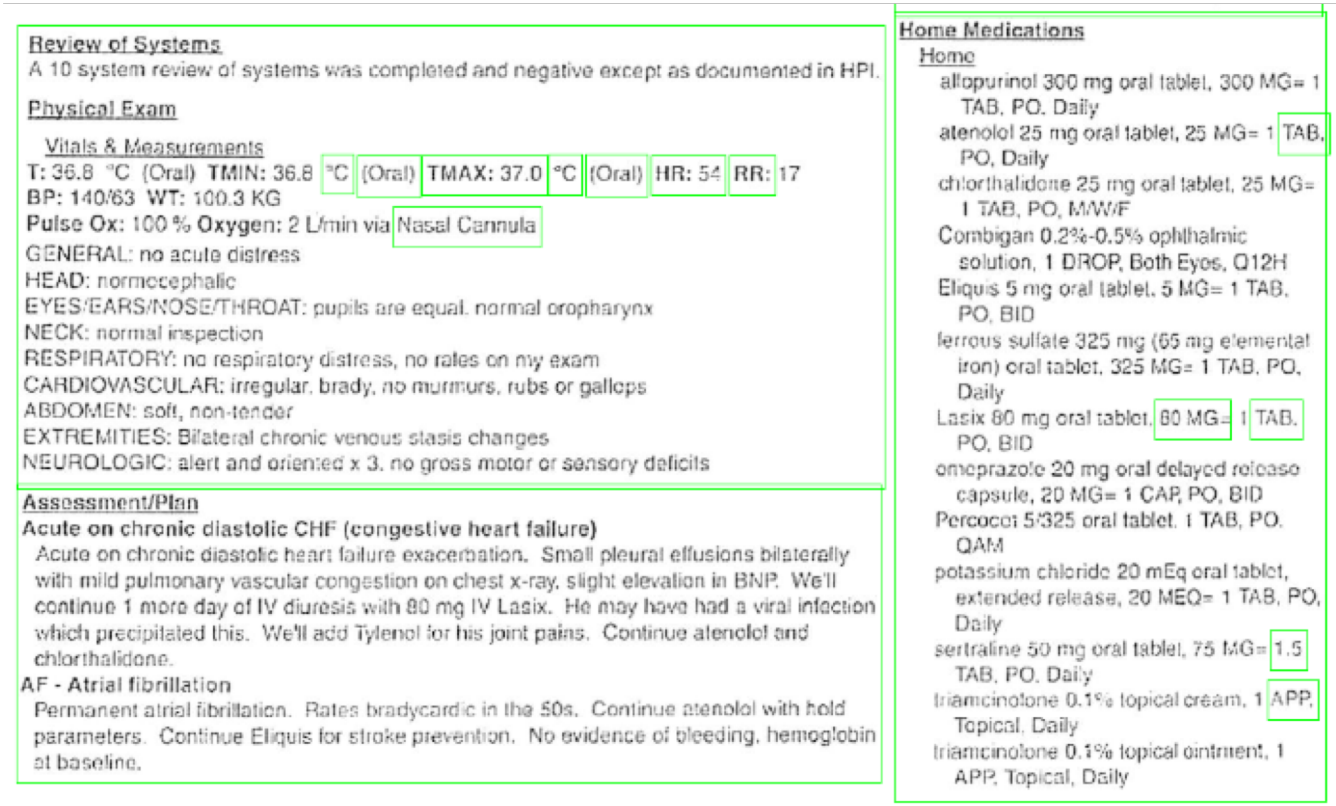
Single execution plan for whole pipeline

OCR CAPABILITIES OF SPARK NLP

OCR in Spark-NLP

- Image preprocessing; scaling, rotation, morphological(erosion, dilation).
- OCR Annotator; layout analysis & text recognition.
    - supports PDFs containing images & text.
    - backed by Tesseract at this point, but open to integrate other engines.
    - works distributed in the cluster, and everything you chain after that.
- OCR Spell Checking for the medical domain.
    - Spell Checking for OCR is special; it depends on print format, paper, font, and language.
    - Error correction method must be adapted to particular domains.

OCR CAPABILITIES OF SPARK NLP - Layout



Review of Systems
A 10 system review of systems was completed and negative except as documented in HPI.

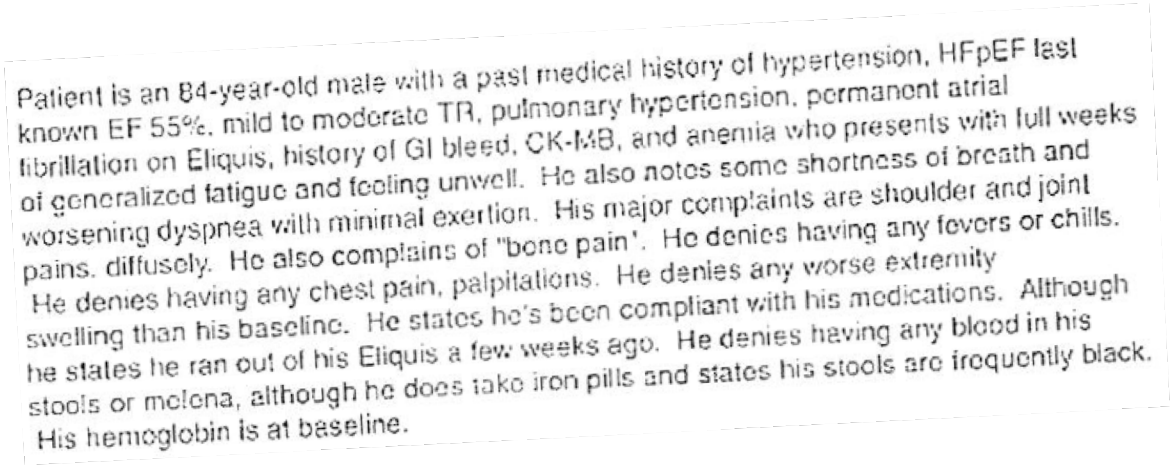
Physical Exam

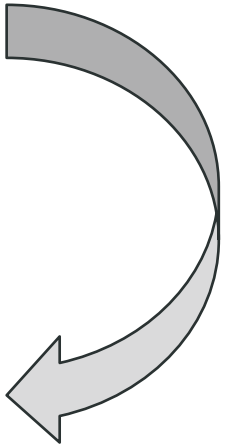
Vitals & Measurements
T: 36.8 °C (Oral) TMIN: 36.8 °C (Oral) TMAX: 37.0 °C (Oral) HR: 54 RR: 17
BP: 140/63 WT: 100.3 KG
Pulse Ox: 100 % Oxygen: 2 L/min via Nasal Cannula
GENERAL: no acute distress
HEAD: normocephalic
EYES/EARS/NOSE/THROAT: pupils are equal, normal oropharynx
NECK: normal inspection
RESPIRATORY: no respiratory distress, no rales on my exam
CARDIOVASCULAR: irregular, brady, no murmurs, rubs or gallops
ABDOMEN: soft, non-tender
EXTREMITIES: Bilateral chronic venous stasis changes
NEUROLOGIC: alert and oriented x 3, no gross motor or sensory deficits

Assessment/Plan
Acute on chronic diastolic CHF (congestive heart failure)
Acute on chronic diastolic heart failure exacerbation. Small pleural effusions bilaterally with mild pulmonary vascular congestion on chest x-ray, slight elevation in BNP. We'll continue 1 more day of IV diuresis with 80 mg IV Lasix. He may have had a viral infection which precipitated this. We'll add Tylenol for his joint pains. Continue atenolol and chlorthalidone.
AF - Atrial fibrillation
Permanent atrial fibrillation. Rates bradycardic in the 50s. Continue atenolol with hold parameters. Continue Eliquis for stroke prevention. No evidence of bleeding, hemoglobin at baseline.

Home Medications
Home
allopurinol 300 mg oral tablet, 300 MG= 1 TAB, PO, Daily
atenolol 25 mg oral tablet, 25 MG= 1 TAB, PO, Daily
chlorthalidone 25 mg oral tablet, 25 MG= 1 TAB, PO, M/W/F
Combigan 0.2%-0.5% ophthalmic solution, 1 DROP, Both Eyes, Q12H
Eliquis 5 mg oral tablet, 5 MG= 1 TAB, PO, BID
ferrous sulfate 325 mg (65 mg elemental iron) oral tablet, 325 MG= 1 TAB, PO, Daily
Lasix 80 mg oral tablet, 80 MG= 1 TAB, PO, BID
omeprazole 20 mg oral delayed release capsule, 20 MG= 1 CAP, PO, BID
Percocet 5/325 oral tablet, 1 TAB, PO, QAM
potassium chloride 20 mEq oral tablet, extended release, 20 MEQ= 1 TAB, PO, Daily
sertraline 50 mg oral tablet, 75 MG= 1.5 TAB, PO, Daily
triamcinolone 0.1% topical cream, 1 APP, Topical, Daily
triamcinolone 0.1% topical ointment, 1 APP, Topical, Daily

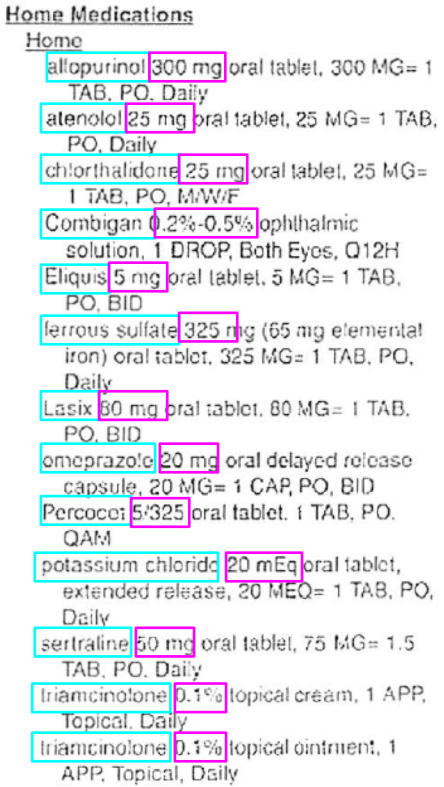
OCR CAPABILITIES OF SPARK NLP - Preprocessing


Patient is an 84-year-old male with a past medical history of hypertension, HFpEF last known EF 55%, mild to moderate TR, pulmonary hypertension, permanent atrial fibrillation on Eliquis, history of GI bleed, CK-MB, and anemia who presents with full weeks of generalized fatigue and feeling unwell. He also notes some shortness of breath and worsening dyspnea with minimal exertion. His major complaints are shoulder and joint pains, diffusely. He also complains of "bone pain". He denies having any fevers or chills. He denies having any chest pain, palpitations. He denies any worse extremity swelling than his baseline. He states he's been compliant with his medications. Although he states he ran out of his Eliquis a few weeks ago. He denies having any blood in his stools or melena, although he does take iron pills and states his stools are frequently black. His hemoglobin is at baseline.



Patient is an 84-year-old male with a past medical history of hypertension, HFpEF last known EF 55%, mild to moderate TR, pulmonary hypertension, permanent atrial fibrillation on Eliquis, history of GI bleed, CK-MB, and anemia who presents with full weeks of generalized fatigue and feeling unwell. He also notes some shortness of breath and worsening dyspnea with minimal exertion. His major complaints are shoulder and joint pains, diffusely. He also complains of "bone pain". He denies having any fevers or chills. He denies having any chest pain, palpitations. He denies any worse extremity swelling than his baseline. He states he's been compliant with his medications. Although he states he ran out of his Eliquis a few weeks ago. He denies having any blood in his stools or melena, although he does take iron pills and states his stools are frequently black. His hemoglobin is at baseline.

Detect sequences



Home Medications

Home

allopurinol 300 mg oral tablet, 300 MG= 1 TAB, PO, Daily

atenolol 25 mg oral tablet, 25 MG= 1 TAB, PO, Daily

chlorthalidone 25 mg oral tablet, 25 MG= 1 TAB, PO, M/W/F

Combigan 0.2%-0.5% ophthalmic solution, 1 DROP, Both Eyes, Q12H

Eliquis 5 mg oral tablet, 5 MG= 1 TAB, PO, BID

ferrous sulfate 325 mg (65 mg elemental iron) oral tablet, 325 MG= 1 TAB, PO, Daily

Lasix 80 mg oral tablet, 80 MG= 1 TAB, PO, BID

omeprazole 20 mg oral delayed release capsule, 20 MG= 1 CAP, PO, BID

Percocet 5/325 oral tablet, 1 TAB, PO, QAM

potassium chloride 20 mEq oral tablet, extended release, 20 MEQ= 1 TAB, PO, Daily

sertraline 50 mg oral tablet, 75 MG= 1.5 TAB, PO, Daily

triamcinolone 0.1% topical cream, 1 APP, Topical, Daily

triamcinolone 0.1% topical ointment, 1 APP, Topical, Daily

medication	**dose**
allopurinol	300mg
atenolol	25mg
chlorthalidone	25mg
combigan	0.2%-0.5%
...	

Code Sample - pipeline

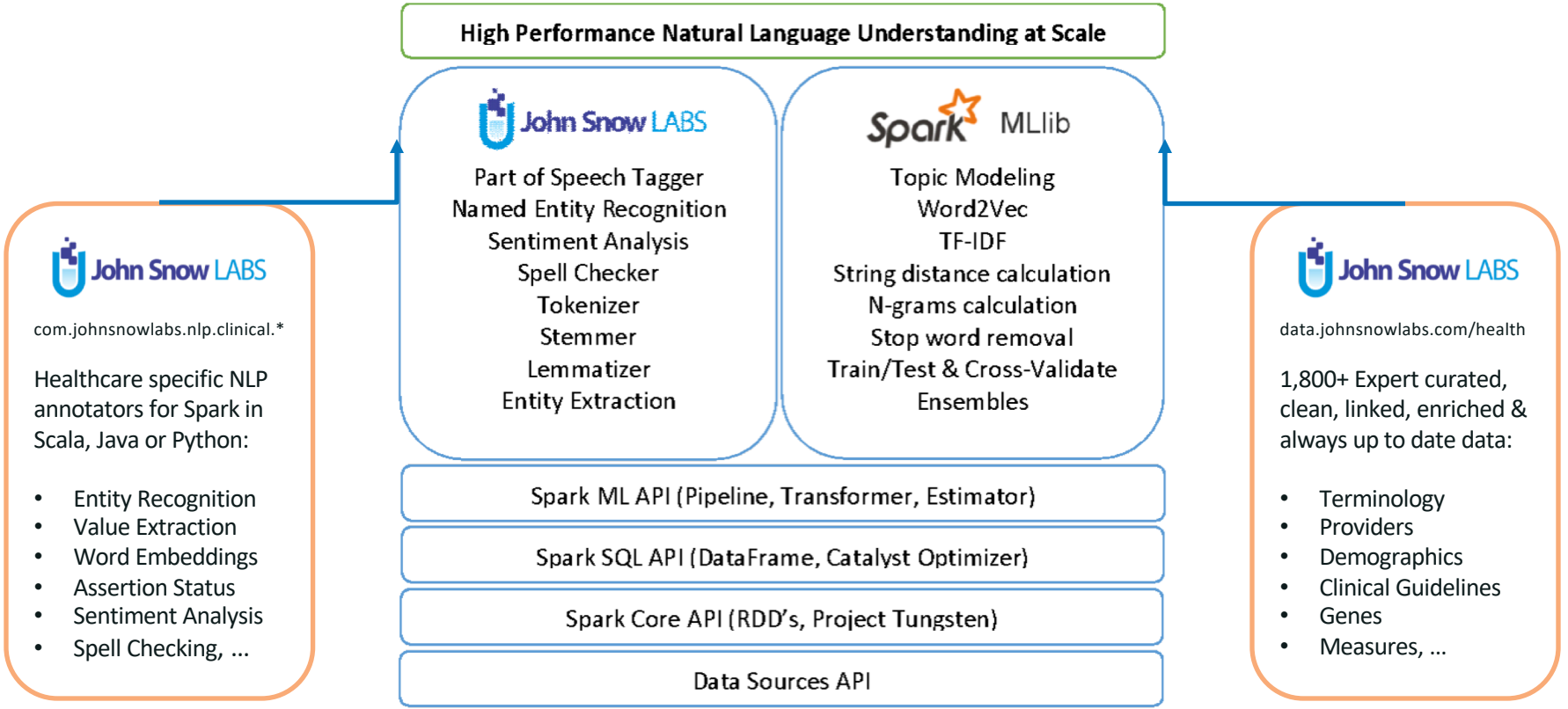
```
OcrHelper.setScalingFactor(2.5f)
OcrHelper.setPageIteratorLevel(PageIteratorLevel.PARAGRAPH)
OcrHelper.useErosion(true, kSize = 1)
val data = OcrHelper.createDataset(spark, inputPath="../strata_demo", outputCol="region",
metadataCol="metadata")

val documentAssembler = new DocumentAssembler().
setInputCol("region").
setMetadataCol("metadata").
setTrimAndClearNewLines(false)
```

```
val token = new Tokenizer().
setTargetPattern("\\S+|\n").
setInputCols(Array("document")).
setOutputCol("token")

val spellChecker = SymmetricDeleteModel.load("schecker").
setInputCols("token").setOutputCol("checked_token")
val finisher = new Finisher().setInputCols("checked_token")

val pipeline = new Pipeline().setStages(Array(documentAssembler, sentenceDetector, token,
spellChecker, finisher))

val result = pipeline.fit(data).transform(data)
result.show()
```

3.

From text to features

SPARK NLP FOR HEALTHCARE



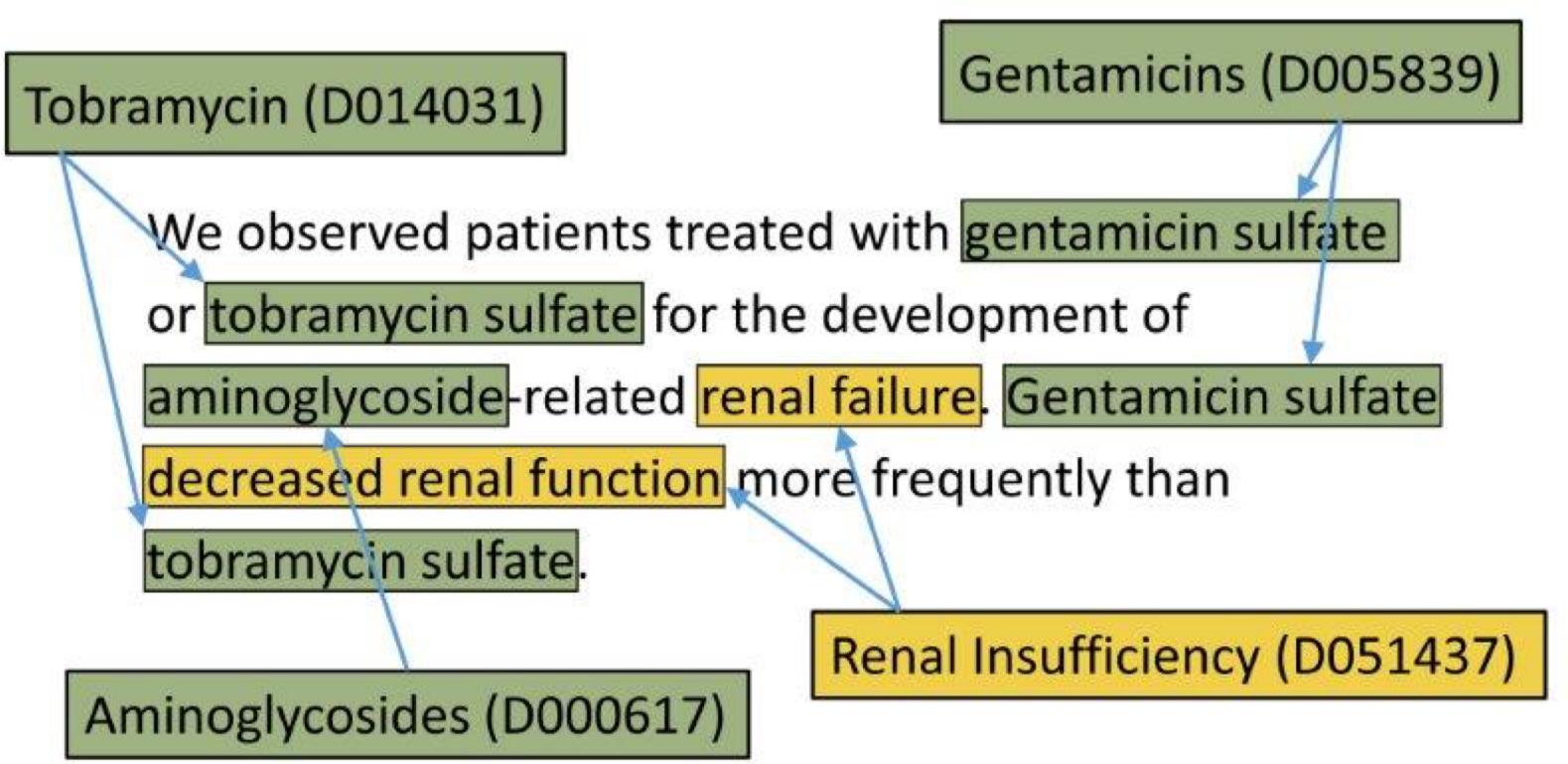
STATE OF THE ART DEEP LEARNING MODELS FOR HEALTHCARE

NLP Task	Implemented Research
Entity Recognition	"Entity Recognition from Clinical Texts via Recurrent Neural Network". Liu et al., *BMC Medical Informatics & Decision Making*, July 2017.
Entity Resolution	"CNN-based ranking for biomedical entity normalization". *Li et al., BMC Bioinformatics, October 2017.*
Word Embeddings	"How to Train Good Word Embeddings for Biomedical NLP". Chiu et al., In *Proceedings of BioNLP'16*, August 2016.
Assertion Status	"Neural Networks For Negation Scope Detection". Fancellu et al., In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016.

NAMED ENTITY RECOGNITION

around the left eye . <test>CT of the brain</test> showed no <problem>acute changes </problem> , <problem>left periorbital soft tissue swelling </problem> . <test> CT of the maxillofacial area</test> showed no <problem>facial bone fracture </problem> . <test> Echocardiogram </test> showed normal left ventricular function , <test>ejection fraction</test> estimated greater than 65% . She was set up with a skilled nursing facility , which took several days to arrange , where she was to be given <treatment>daily physical therapy</treatment> and <treatment> rehabilitation </treatment> until appropriate .

ENTITY NORMALIZATION



ASSERTION STATUS DETECTION

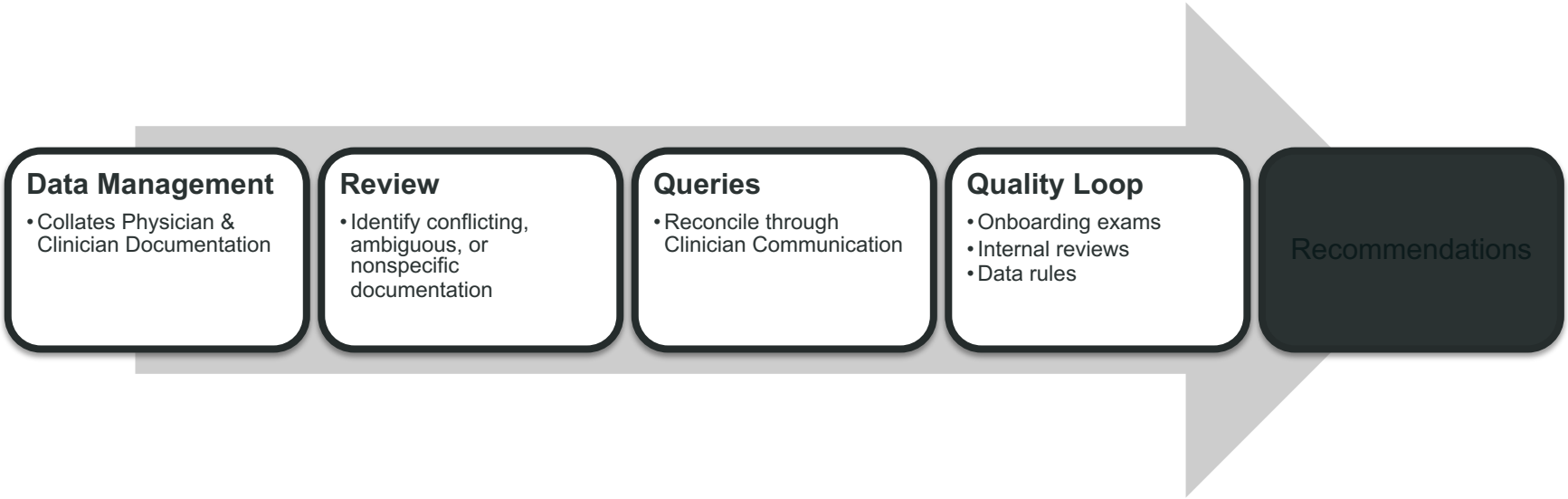
Prescribing sick days due to diagnosis of **influenza**.	*Present*
Jane complains about **flu-like** symptoms.	*Possible*
Jane's **RIDT** came back clean.	*Absent*
Jane is at risk for **flu** if she's not vaccinated.	*Hypothetical*
Jane's older brother had the **flu** last month.	*Family history*
Jane had a severe case of **flu** last year.	*Patient history*

4.

From features to models

Silver Tsunami, Regina the Reviewer and Payment Models

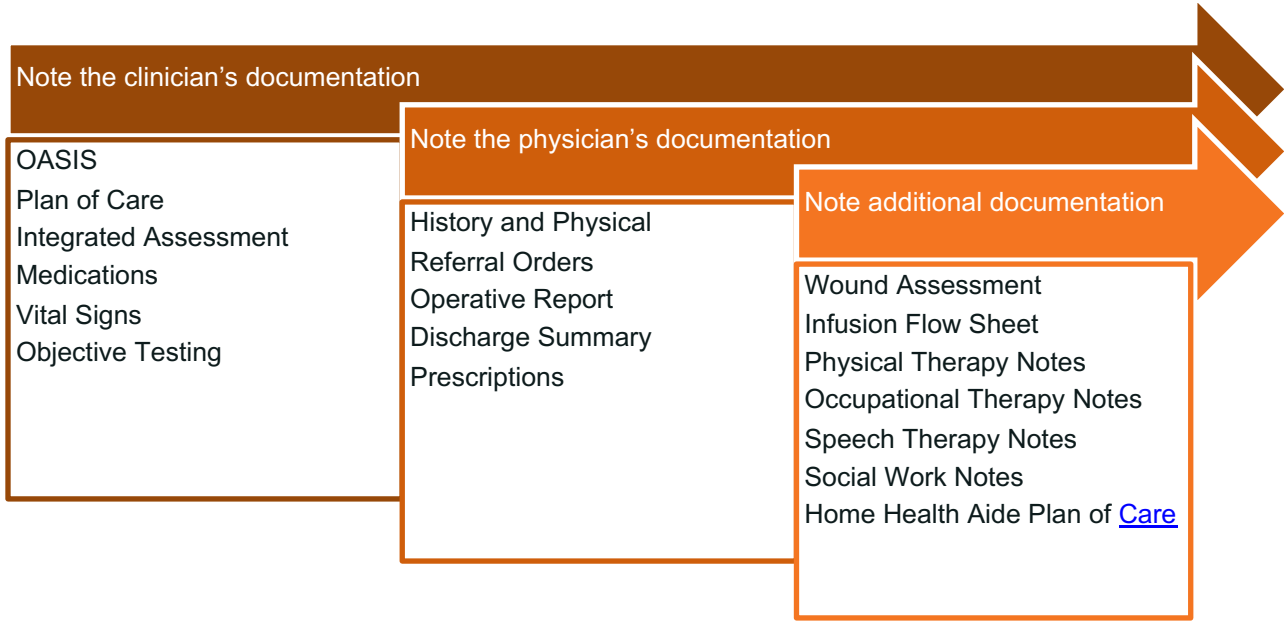
The Current Process



GAPS

- Data Management
  - We currently have Data Processing Operators that fulfill this function. Moving documents from one system to another.
  - Least Expensive-Readily Available-EMRs don't play well together
- The Coding Process
  - Most Expensive, most experienced resource
  - Area with the widest gaps
  - Gap between experienced and inexperienced = Abstracting the record (identifying what is important)
- Queries
  - Not Standardized-Relies on Decision Making at the Coding Specialist Level
  - Experience…

The Current Process



The Current Process

Cupcakes, Puppies and Bombs

User Stories:
As a Manager I want to be able to identify assessments that are Hard, Medium and Easy by two metrics, degree of effort and perceived level of difficulty.

As a Manager I want to be able to identify employees that possess the appropriate skill level to complete the work for an assessment that has been categorized as Hard, Medium and Easy

As an Employee I want to be appropriately matched to a work assignment based on my ability to effectively complete that work assignment.

Cupcakes, Puppies and Bombs

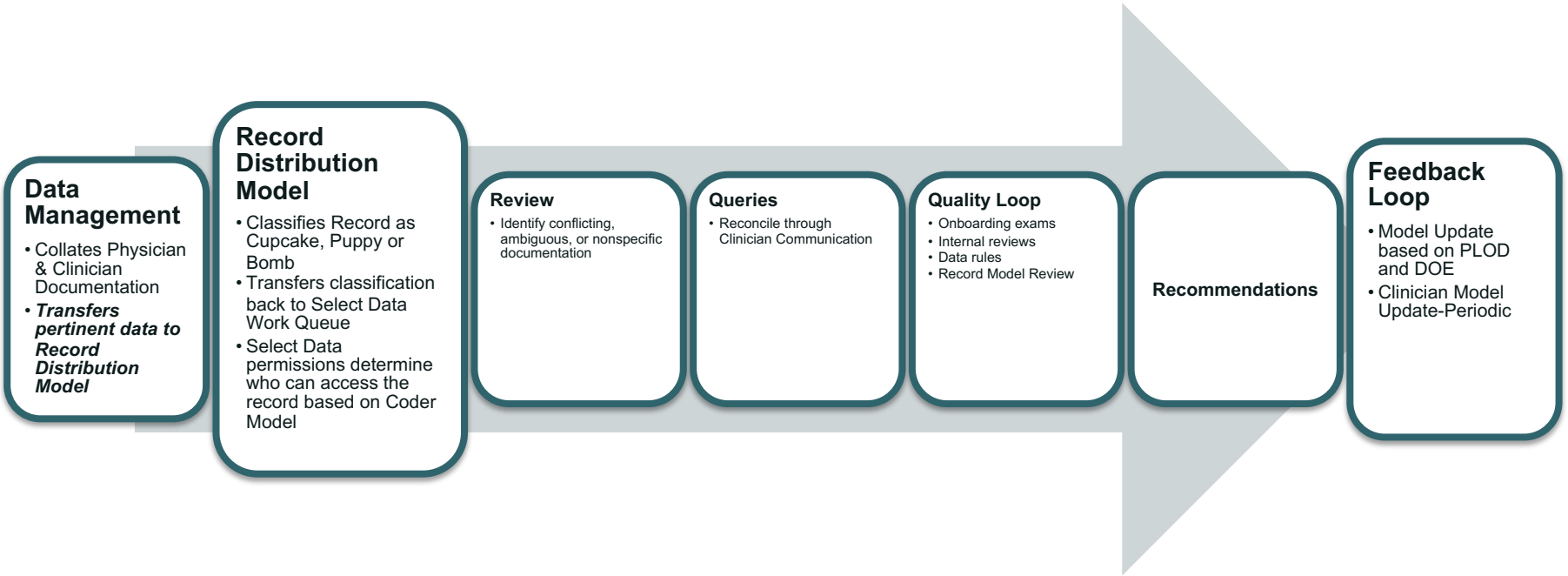
Cupcakes, Puppies and Bombs

- Research shows that production increases if you perform your most difficult task at the beginning of the day
- Coding Specialists have felt a disproportionate distribution of records
- Users want a way to measure record difficulty and match it with the employee's ability to accurately and productively complete that record

Our Hypothesis:

We can increase our overall production by 10% and maintain an accuracy of 95% with our current staff by instituting a workflow that identified the records based on perceived level of difficulty and matched those records to individuals who are competent to complete the records accurately and productively.

Cupcakes, Puppies & Bombs



Cupcakes, Puppies & Bombs

- Feedback Loop
  - Perceived Level of Difficulty-PLOD
    - Subjective measurement
    - Validated using objective measures
    - Comparison among levels of experience among employees
    - Normalized to account for these differences.
  - Degree of Effort-DOE
    - Measured using minutes spent within the record
    - Time-stamps from all applications transformed into minutes
  - Model Reliability
    - Text that was not identified
    - Text identified incorrectly
    - Text identified correctly-Context not identified appropriately

From Cupcakes to Automated Coding

- Distrust of AI among healthcare professionals
  - Emphasis placed on intuition
  - General lack of knowledge regarding programming
  - Fear of lost employment
- Distrust to Trust
  - Exposure
  - Begin by replacing small pieces of the mundane within the process
  - Augmented Intelligence
  - Gradually transform the role
  - Coding Specialist then becomes a Quality Specialist – focused on ensuring the accuracy of the model

THANK YOU!

To try Spark NLP:

https://nlp.johnsnowlabs.com

Getting Started, Documentation,
Examples, Videos, Blogs, Code,
and an active Slack Community

To bounce ideas:

Alberto Andreotti
Stacy Ashworth
Tawny Nichols
David Talby

John Snow LABS

SelectData™