NEC *Relentless* passion for innovation

Retrieving Medical Records with SENNAMED: NEC Labs America at TREC 2012 Medical Record Track

Yanjun Qi and Pierre-Francois Laquerre

Department of Machine Learning, NEC Laboratories America Princeton, New Jersey, USA

yanjun@nec-labs.com, pierre.francois@nec-labs.com

NECLA "sennaMed" System

- Three text analytic modules (we believe) are necessary for "Semantic Retrieval of Medical Records"

- (I) Medical report understanding
  - Medical concept extraction and categorization
    - "knee surgery" is a procedure
    - "nausea" is a symptom
  - Report information extraction
    - Extract implicit information, e.g. the patient's gender even if "male" or "female" are not explicitly mentioned
    - Tell the difference between the patient's history, her family history, her current state

- (II) Query text understanding
  - Similar to "report understanding" as above

- (III) Classic IR retrieval → query vs. medical records

sennaMed (I): Report understanding

We need to interpret the reports into meaningful medical profiles

➔ Term Transformation to UMLS Semantic Concepts

A 46-year-old lady was referred to the haematology clinic for evaluation of lymphocytosis in May 1993. She had severe lethargy and intermittent right upper abdominal discomfort with significant loss in weight. Her past medical history included essential hypertension controlled on atenolol 100 mg once daily and was also on frusemide 40 mg once daily. She had no significant *surgical history* other than having undergone cholecystectomy in 1972. She is not a *smoker*. Physical examination showed no evidence of *lymphadenopathy*. Peripheral blood showed numerous Howell-Jolly bodies within erythrocytes. Thyroid function tests, protein electrophoresis, C-reactive protein, immunoglobulin levels and autoimmune screening were normal. Ultrasonography and computed tomography scan of the abdomen and pelvis did not reveal mediastinal lymphadenopathy, but the spleen was noted to be very atrophic. Gastroscopy showed multiple gastric erosions and the initial impression was of celiac disease complicated by lymphoma and lymphocytosis.

Categories

- Profile
- Symptoms
- Medication
- Treatments
- Diseases
- Tests
- Others

sennaMed (II): Query understanding

Similarly, we need to interpret the query topic

1) Women with hearing loss

2) Patients admitted for care who take herbal products for osteoarthritis

3) Diabetic patients who received diabetic education in the hospital

4) Cancer patients with liver metastasis treated in the hospital who underwent a procedure

5) Patients with GERD who had esophageal adenocarcinoma diagnosed by endoscopy

Categories

Profile
Symptoms
Medication
Treatments
Diseases
Tests
Others

sennaMed (III): Retrieval Modules

- Two classic retrieval models for indexing and ranking
  - (1) a vector space retrieval model and
  - (2) a language model based retrieval approach.
- Also test several other classic IR techniques, which include
  - (1) dimension reduction using latent semantic indexing, and
  - (2) query expansions.

Our four submissions

NEC Laboratories America
Relentless passion for innovation

Official Runs	Description
sennamed1	UMLS concept representation, language model retrieval, query expansion
sennamed2	UMLS concept representation, vector space retrieval, query expansion
sennamed3	UMLS concept representation, vector space retrieval
sennamedlsi	UMLS concept based representation, vector space retrieval, LSI

Parameters are tuned on the 2011 TREMed test data.

Also tried other methods (SSI, LDA, query expansion variations, etc.)

Metamap: medical concepts extracted as words, with negated concepts included as separate terms (e.g. “smoker” is different than “not a smoker”)

NEC Laboratories America
Relentless passion for innovation

Global results

Metric	sennamed1	sennamed2	sennamed3	sennamed1LSI
infAP	0.2246	**0.2745**	0.2169	0.2151
infNDCG	0.478	**0.5468**	0.4688	0.4468
P@10	0.5255	**0.5574**	0.5447	0.4468
R-prec	0.3457	**0.3805**	0.3298	0.2974
bpref	0.3647	**0.427**	0.3559	0.3496
recip_rank	0.7706	**0.7696**	0.7497	0.6189

sennaMed2 run (3rd overall / 2nd in automatic runs)

NEC Laboratories America
Relentless passion for innovation

# topics	infAP	infNDCG	R-prec	P@10
Best	5	6	5	11
Better than median	27	27	24	13
Median	2	1	6	12
Below median	13	13	10	7
Worst	0	0	2	4
Summary (good/ bad)	34 / 13	34 / 13	35 / 12	36 / 11

EXTRA about Concept Detection Using MetaMap

Courtesy slides from
MetaMap 2010 Tutorial

Why Concept Identification?

- Structured data vs. text
- Concept identification is useful/essential for many tasks including
  - Information extraction/Data mining
  - Classification/Categorization
  - Text summarization
  - Question answering
  - Literature-based knowledge discovery

Courtesy slides from
MetaMap 2010 Tutorial

Example (best mappings ***with WSD***)

- PMID – 9339686
- AB –Cerebral blood flow (CBF) in newborn infants is

Cerebrovascular Circulation

Infant, Newborn

~~CEREBRAL BLOOD FLOW IMAGING~~

often below levels necessary to sustain brain viability

Frequent

Levels (qualifier value)

Sustained

Brain

Viable

~~Entire brain~~

in adults.

Adult

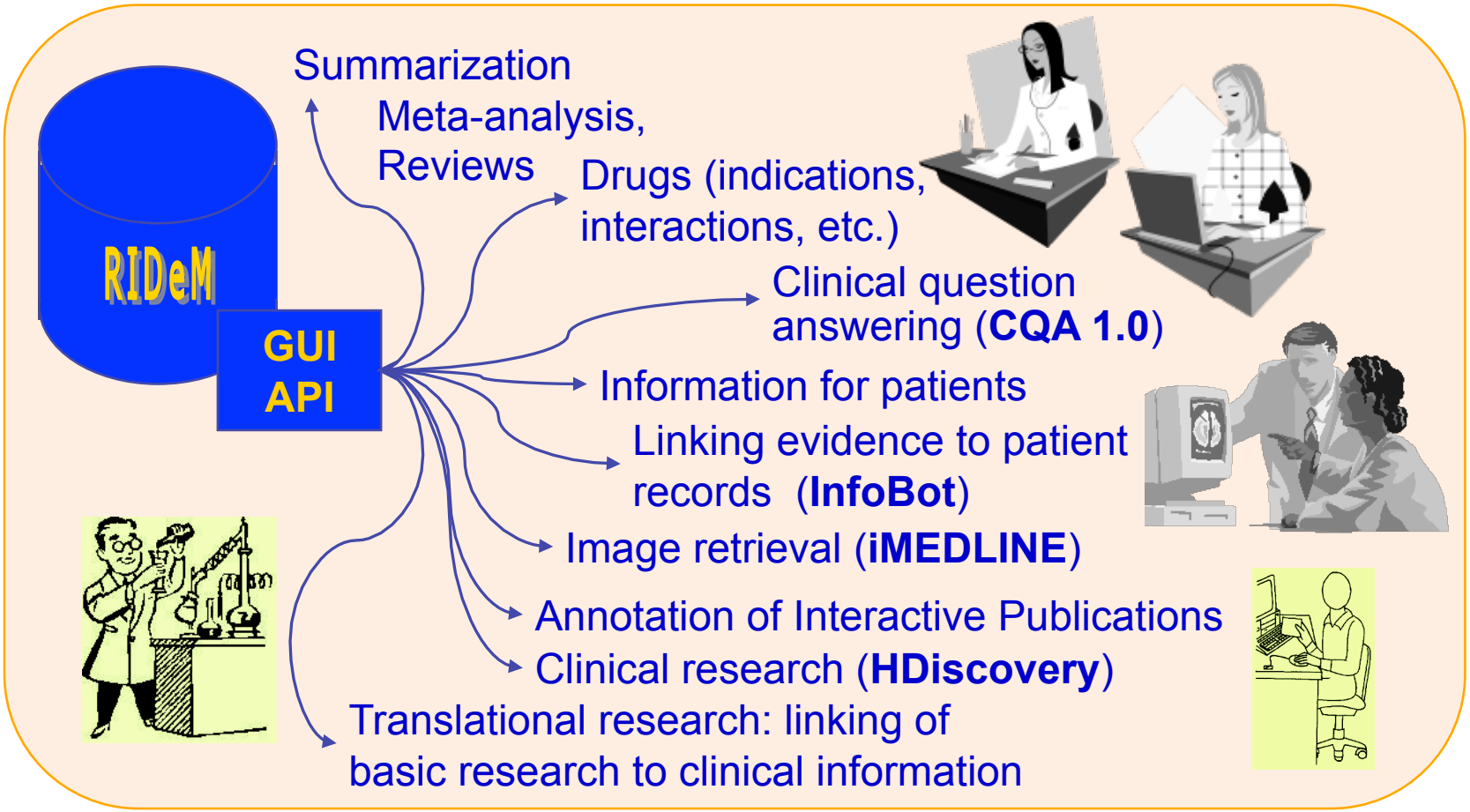
Courtesy slides from
MetaMap 2010 Tutorial

The Algorithm

- Parsing
  - Using SPECIALIST minimal commitment parser, SPECIALIST lexicon, MedPost part of speech tagger
- Variant generation
  - Using SPECIALIST lexicon, Lexical Variant Generation (LVG)
- Candidate retrieval
  - From the Metathesaurus
- Candidate evaluation
- Mapping construction

Courtesy slides from
MetaMap 2010 Tutorial

NLM Applications using MetaMap: RIDeM (Repository for Informed Decision Making)



Courtesy slides from
MetaMap 2010 Tutorial