SNOMED CT EXPO 2020
Virtual Conference October 8-9

Entity tagging for SNOMED CT concepts in text-based telemedical solutions

Pawel Paczuski, CEO @ upmedic - startup looking for opportunities related to SNOMED CT implementations
Artur Kusak, Medical Centers the Medici, Lodz (Poland) - networks of clinics with more than 20 years of experience

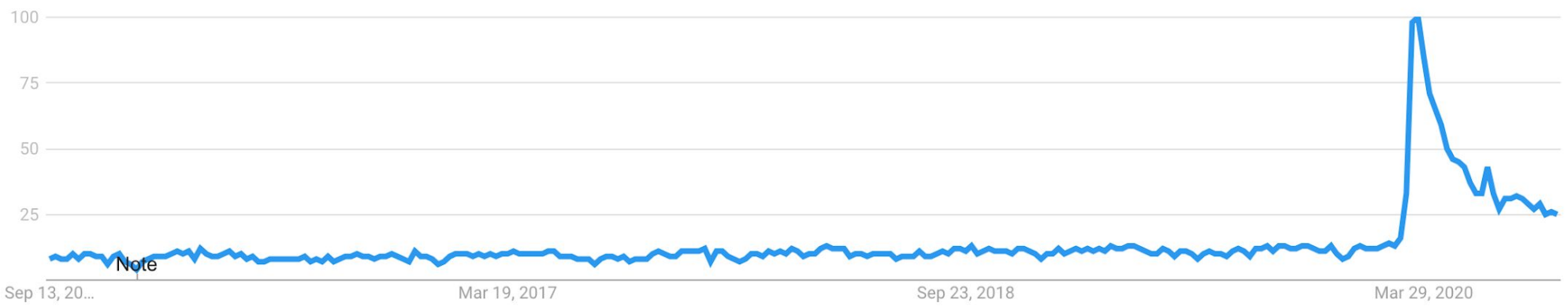
COVID-19 and telemedicine

telemedicine
Search term

+ Compare

Worldwide | Past 5 years | All categories | Web Search

Interest over time

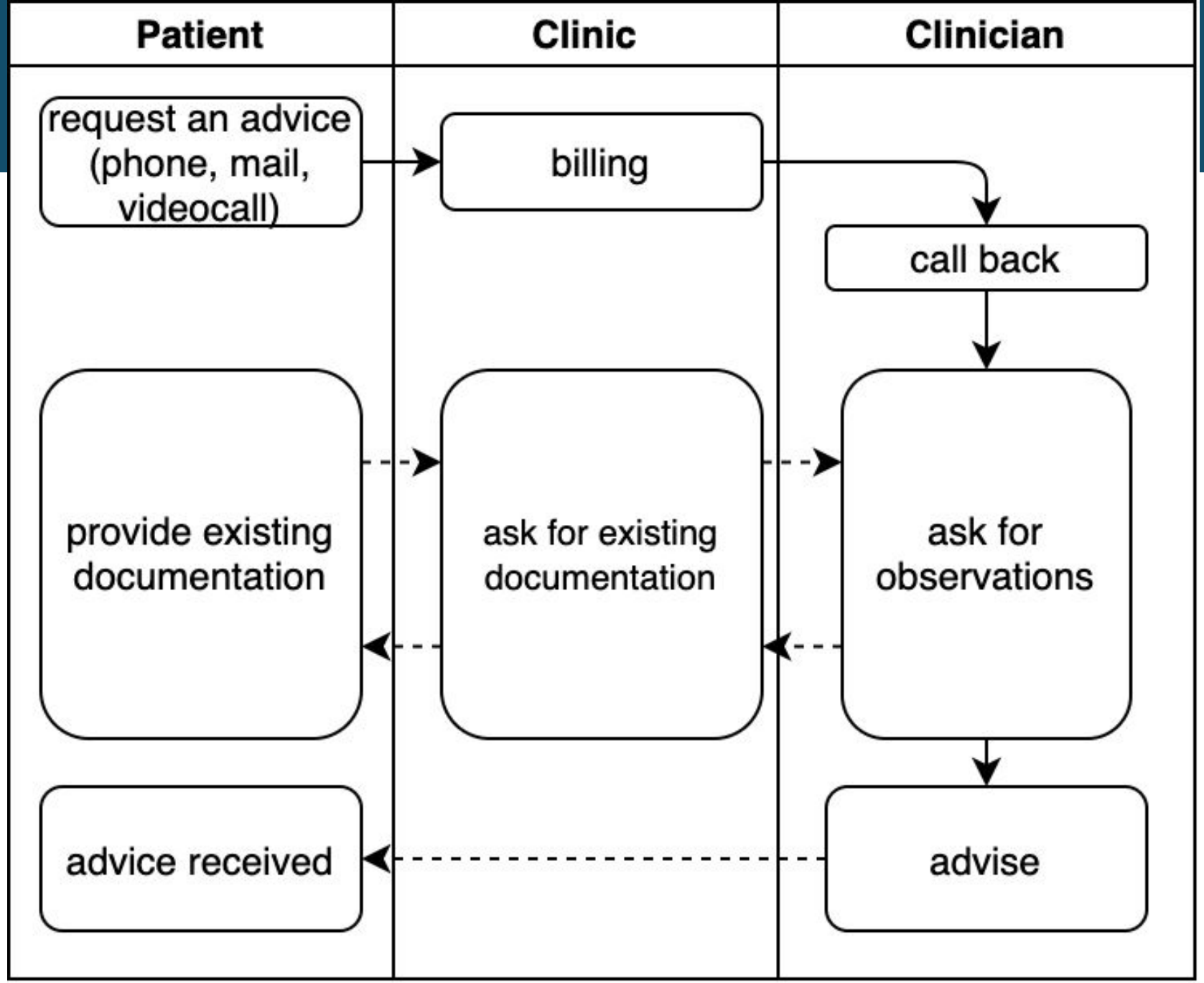


COVID-19 and telemedicine continued

- People are more aware of telemedical solutions and their possibilities now than ever
- Sometimes it is the only available way of communicating with clinicians without risking anyone’s life
- Good for non-emergency situations
- Especially suitable for teleradiological advice, however, not most popular

Teleradiological advice

Most popular workflow of a teleradiological advice


Patient
Clinic
Clinician
request an advice (phone, mail, videocall)
billing
call back
provide existing documentation
ask for existing documentation
ask for observations
advice received
advise

Typical teleradiological advice information flow

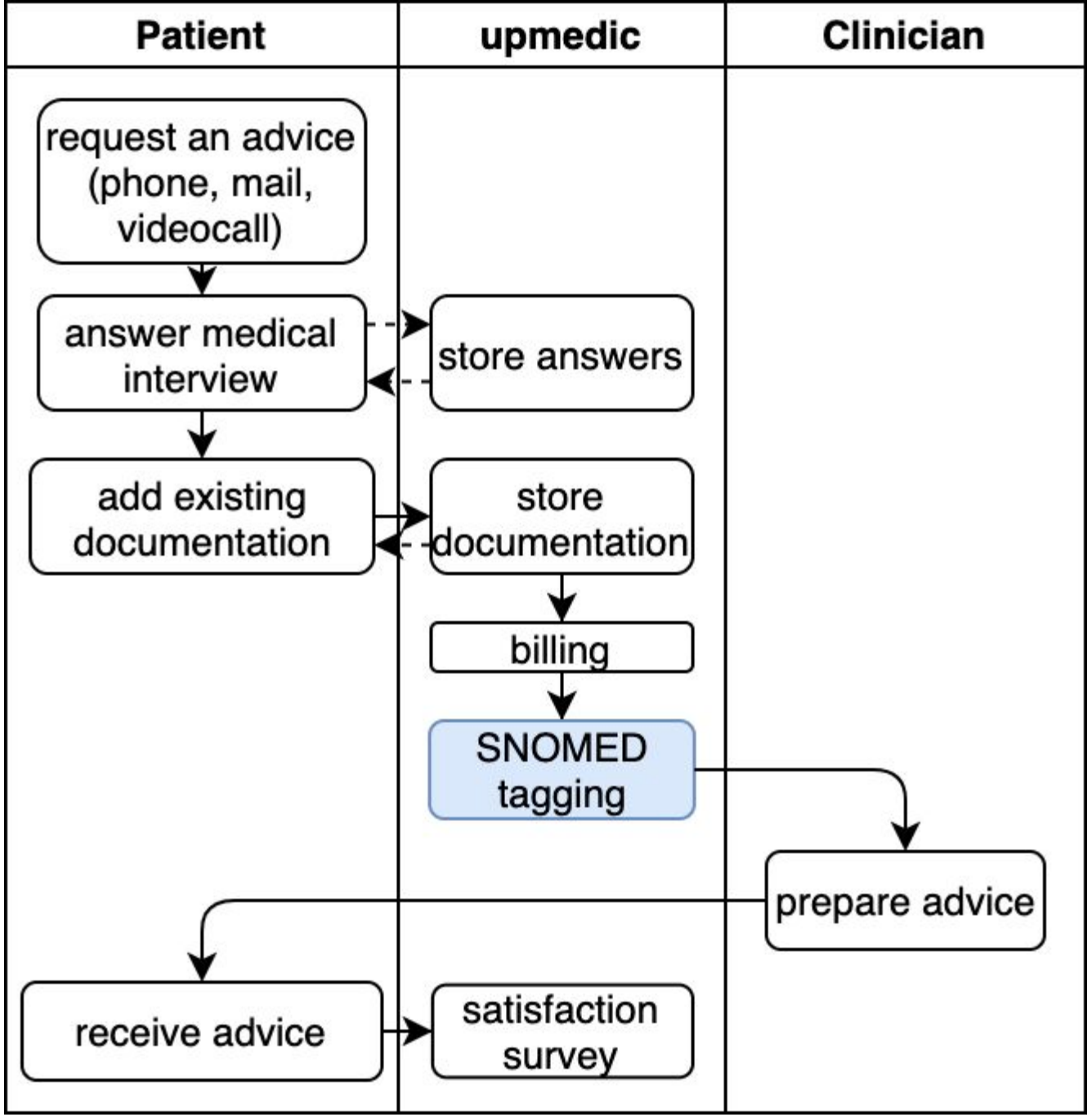
Good

1. Similar to the traditional appointment with a clinician
2. Easy to make an advice for untypical cases because clinician can ask patient any question

Bad

1. Synchronous (both patient and clinician are engaged simultaneously during the phone/video call)
2. Basis for diagnosis contained in recordings of the call, mailing, etc.
3. Difficulties with exchanging existing medical documentation

Proposed workflow of a telemedical advice


Patient
upmedic
Clinician
request an advice (phone, mail, videocall)
answer medical interview
store answers
add existing documentation
store documentation
billing
SNOMED tagging
prepare advice
receive advice
satisfaction survey

Asynchronous teleradiological advice

Good

1. Clinicians are no longer responsible for conducting interview for typical cases (30% of time of an average appointment)
2. Asynchronous (patients might make a request for advice anytime, clinicians answer when available)
3. Patients' observations become a part of the documentation automatically
4. SNOMED CT attached to patient's observations

Bad

1. Requires falling back to the traditional information flow for rare, untypical cases not covered by the interview
2. Learning curve attached

upmedic

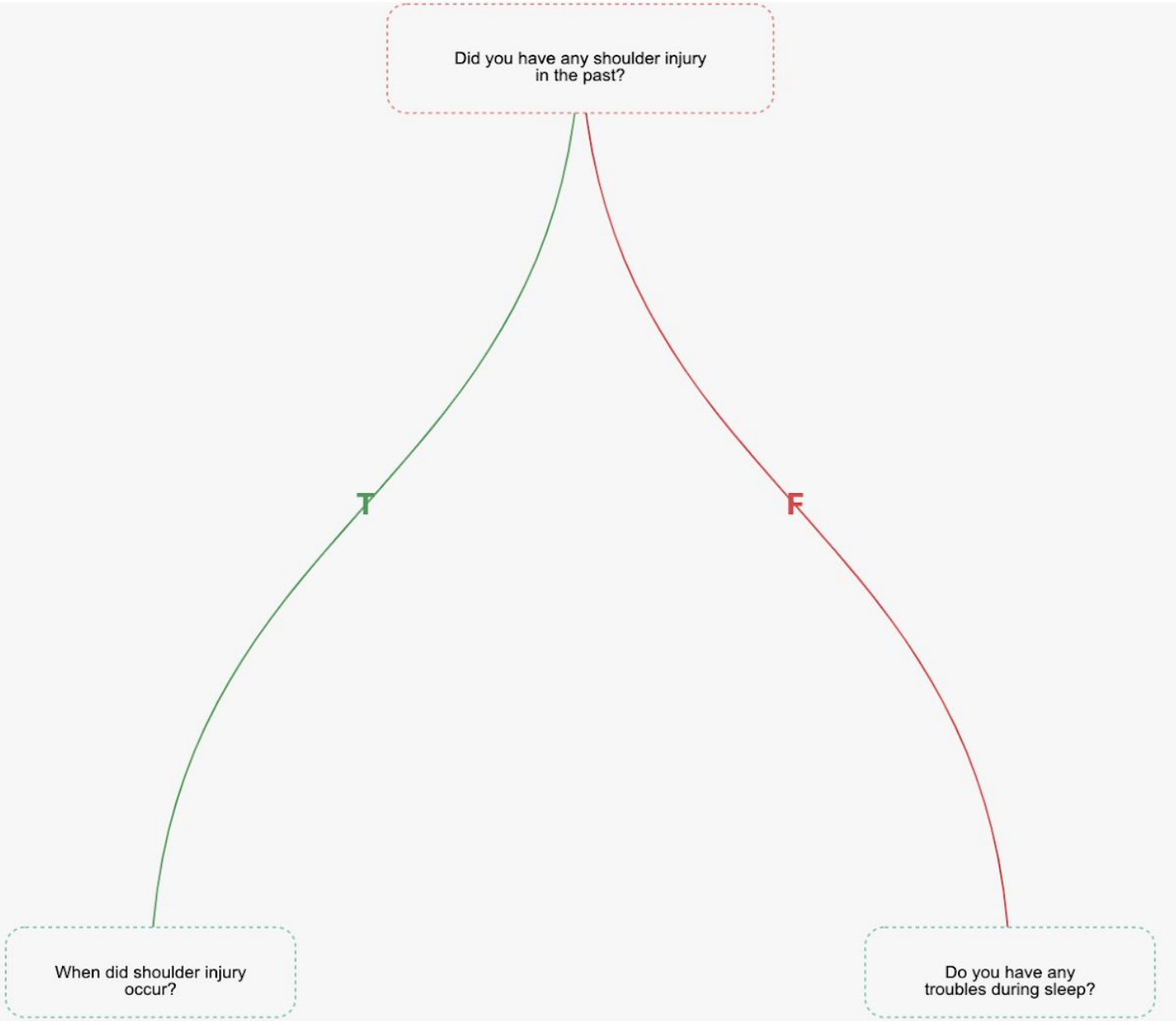
Survey's name:

Before shoulder examination

List of questions

- Previous shoulder examinations (none, ultrasonography, Radiography, CT, MRI)
- Did you have any shoulder injury in the past?
- Do you feel any pain?
- Under what circumstances does the pain appear?
- Can you lay on the aching side?
- Does the pain interrupt sleep?

Question tree



Why should we attach SNOMED CT to patients' observations?

- AI-driven triage for patients
- Automatic mapping SNOMED CT to ICD-10
- Lie detector using SNOMED CT
- Auditing

How to attach SNOMED CT to plaintext phrases?

- observe nature of the problem
  - medical concepts (fuzzy-matched with SNOMED CT concepts)
  - contexts (sets of phrases between two consecutive medical concepts)
  - phrases (pieces of text that might be fuzzy-matched with SNOMED CT concepts)
- entity tagging to find medical concepts
- fuzzy matching with SNOMED CT descriptions to match as many SNOMED CT concepts as possible

Plaintext radiological reports form

computed tomography of the chest examination technique: examination from left-sided administration. subclavian - only contrast is visible in the brachocephalic orifice and has not reached the parenchymal organs of the chest and abdominal cavity, trace in a delayed examination . right lung: airless, surrounded by fluid; it is not possible to assess tumor mass in a test without cm. . left lung: correctly airborne ; no densities ; without focal lengths . lymph nodes in the mediastinum: not enlarged . trachea and bronchi: infiltration and obstruction of the left bronchus . pleural cavity: the fluid in the right cavity fills the entire cavity and compresses the inflamed lung .

Data and tooling

- 100 labelled radiological reports machine translated into English containing around 3000 medical concepts and 3000 contexts
- SpaCy with English language model for deep named entity tagging (Bloom embeddings) to find medical concepts, natural language processing, sentence splitting, noun-chunks splitting
- PyMedTermino - medical ontologies exploration api used for SNOMED CT relational-modelling for SNOMED CT RF2 format and efficient querying SQLite database that stores SNOMED CT data
- fuzzysearch for... fuzzy matching!

Named entity recognition

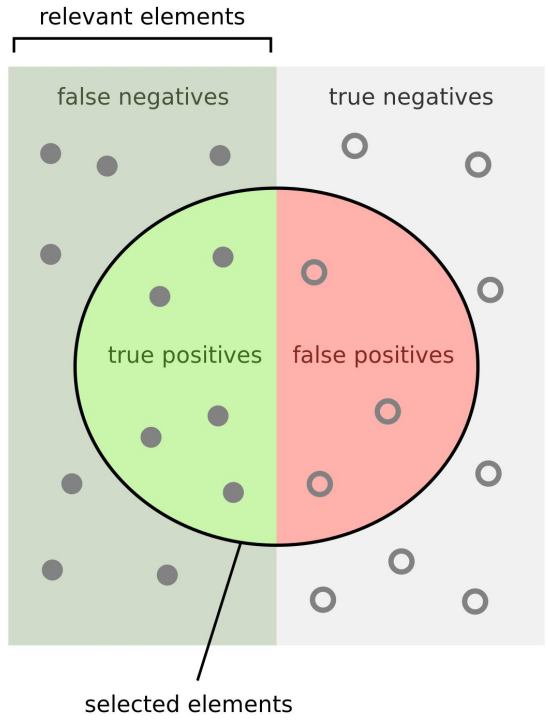
- Transfer learning using English language model created using Bloom embeddings
- Catastrophic forgetting problem - solved using pseudo-rehearsing
- For testing set input: tag MEDCONCEPTs in free-text - root phrases that can me mapped to SNOMED CT concept and are then used for querying SNOMED CT for phrases within concepts

Contexts splitting

- Mark pieces of text between two neighboring MEDCONCEPTs as contexts
- Split contexts into noun-chunks generating phrases
- Match phrases with SNOMED CT concepts (stemming, fuzzy matching included, a lot of experimenting) starting with the longests ones - greedy chunking makes it more probable to find (if exists) a SNOMED CT concept that is more semantically similar to the phrase that is being considered

Measures of success of tagging

- Precision
- Recall
- F1-score - harmonic mean of precision and recall



How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{selected elements}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{relevant elements}}$$

Our results versus others

- J. Patrick, Y. Wang, and P. Budd, “An automated system for conversion of clinical notes into snomed clinical terminology” - dictionary matching
- I. M. Soriano and J. Castro, “Dner clinical (named entity recognition) from free clinical text to snomed-ct concept” results are "promising"
- A. Arbabi, D. R. Adams, S. Fidler, and M. Brudno, “Identifying clinical terms in free-text notes using ontology-guided machine learning” (precision, recall, F1-score) = (79.5%, 62.1%, 69.7%)

- Our results : (precision, recall, F1-score) = (75.5%, 84.4%, 79.1%)

Conclusions

- Tagging SNOMED CT concepts is a difficult problem but there exists open-source tooling that makes it easier to load and process SNOMED CT
- Loaded SNOMED CT can be then processed using general-purpose NLP tools
- Having SNOMED CT concepts attached to the free-text documentation, even if not perfect, can be a starting point for interesting solutions, e.g. automatic triage, patients' lie detector, auditing, etc.

References

- Bodenreider, O., Smith, B., Kumar, A., Burgun A. (2007) "Investigating subsumption in SNOMED CT: An exploration into large description logic-based biomedical terminologies", Artificial Intelligence in Medicine.
- Sacks, H. (2015) "Knowledge Engineering with Semantic Web Technologies". Online: https://www.youtube.com/watch?v=eJ9H1SakPoA
- Paczuski, P. (2020), “Natural Language Processing system for automatic structuring medical free text documentation into SNOMED CT ontology concepts”, Warsaw University of Technology.

Contact

Eager to have any discussions with people interested in solutions using SNOMED CT

Contact: pawel.paczuski@upmedic.io