ALMA MATER STUDIORUM UNIVERSITY OF BOLOGNA
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING – DISI

Neuro-Symbolic AI for Natural Language Understanding from Text in Health Domains

Giacomo Frisoni, Gianluca Moro

DISI – University of Bologna, Cesena
Via dell'Università, 50 I-47522 Cesena (FC), Italy
giacomo.frisoni@unibo.it, gianluca.moro@unibo.it

April 27th, 2022

What’s the context?

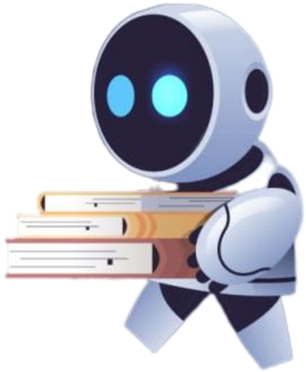
Natural Language Understanding

" ***Language is at the heart of human intelligence****. It therefore is and must be at the heart of our efforts to build artificial intelligence.* ***No sophisticated AI can exist without mastery of language****.*

- Rob Towes, Language Is The Next Great Frontier In AI, Forbes

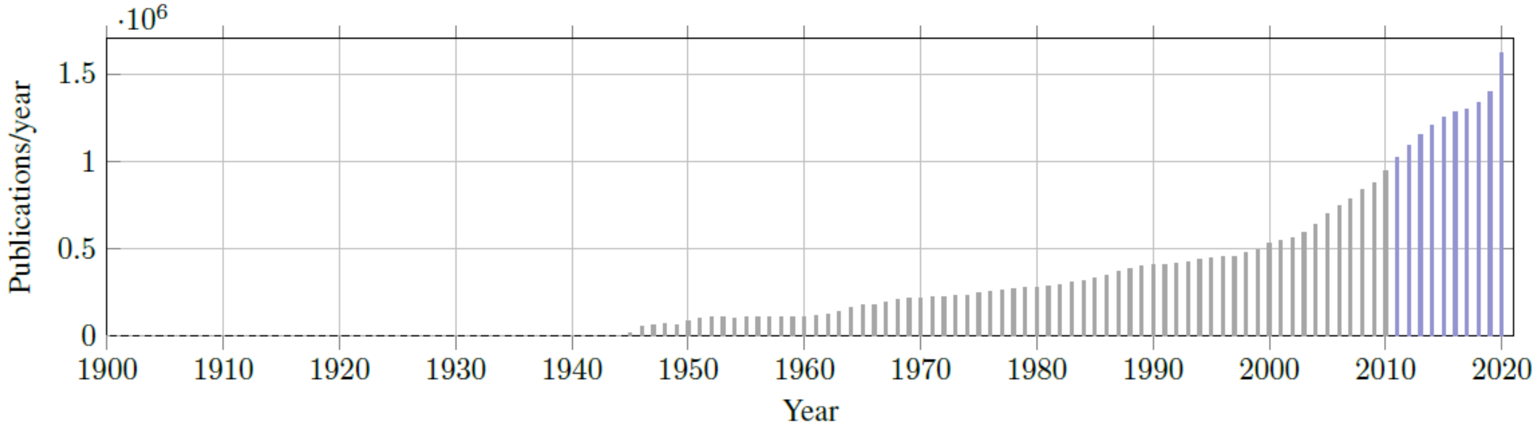
" Before, if an AI system could translate a text, it would have been deemed as having understood the text, or if the system could sustain a conversation or describe the scene on an image. ***Now, none of these count as proper understanding****.*

- Geoff Hinton, Turing Award 2018



Exponential growth of text documents

- High and growing availability of unstructured textual content on the Web
  - Humans are unable to consume the information contained in text documents of their interest at the same rate as they are produced and accumulated over time
  - Importance of natural language understanding (NLU) and text mining to automatically extract useful knowledge from the text
- The scientific literature has led to an information overload
  - Biology is feeling most of this literature congestion blow, witnessing exponential growth in in publications over the last few decades



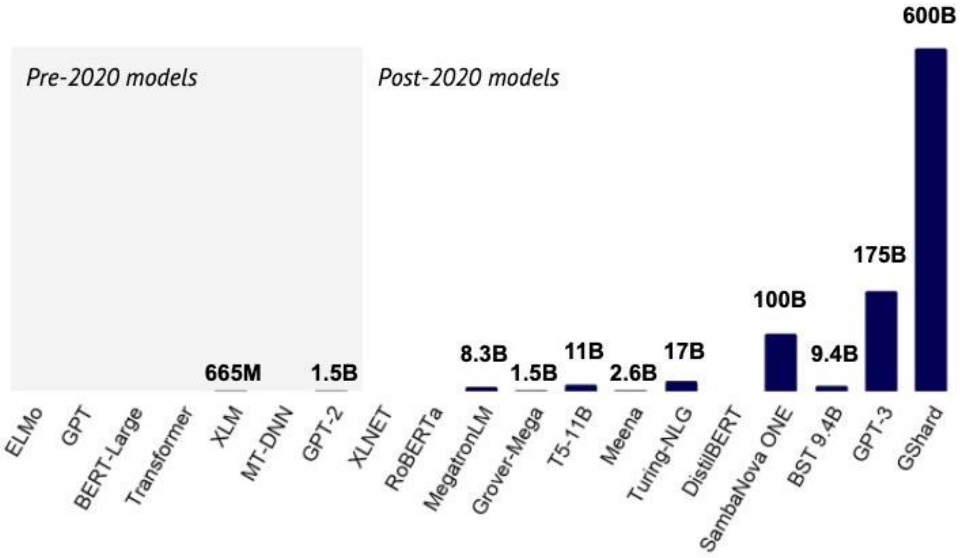
  - More than 1 million new records pour into the PubMed database each year
  - From 2020, about ***3 papers per minute!***

Language Models

- Language models (LMs) are a key component of modern NLP
  - LMs represent probability distributions over a vocabulary, eventually conditioned on a context, like a window of words or external knowledge
  - Neural LMs are trained as probabilistic classifiers over continuous and low-dimensional representations of words, called *word embeddings*
  - Modern continuous space embeddings encapsulate the **contextualized meaning** of tokens, disambiguating polysemes and quantifying semantic similarities
  - Pre-training on extreme-scale unlabeled text corpora has shown to capture a **surprising amount of linguistic and relational knowledge stored in model parameters**
  - Since the introduction of the Transformer architecture, we are witnessing an unbridled release of new models (e.g., BERT, T5, GPT-2) and their adaption to downstream tasks
  - **Many contributions in the medical field** (e.g., SciBERT, BioBERT, PubMedBERT, ClinicalBERT, Med-BERT)
- Before neural solutions…
  - Algebraic LMs (e.g., LSA on BOW + TF-IDF), Probabilistic LMs (e.g., pLSA, LDA)
- State-of-the-art results in many tasks related to the healthcare domain
  - Medical QA, Multi-document summarization, Medication recommendation, Phenomena explanation, Gene function finding

Open Problems – i

- Latest advances have highlighted several weakness of pre-trained models
  - **Black-box knowledge** due to high-dimensional parametric spaces (continuously growing) which make almost impossible to demonstrate what a model has really learned
  - Reached complexities require computational resources to now only accessible to large multinationals, undermining **research democracy**
  - Many models can produce **hallucinations** (i.e., not factually accurate), are **fragile** (i.e., can easily be fooled), **biased**, and **toxic** (e.g., gender inequality and hate speech)
  - Without **explainability** and without verifying the knowledge learned, the new progress of deep learning risks not enjoying the **trust** of users anymore and causing a new AI crisis



Gopher from DeepMind
Less than 48 hours ago
280B

Megatron-Turing NLG from Nvidia
One month ago
530B

Open Problems – ii

On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜

Emily M. Bender*
ebender@uw.edu
University of Washington
Seattle, WA, USA

Timnit Gebru*
timnit@blackinai.org
Black in AI
Palo Alto, CA, USA

Shmargaret Shmitchell
hmargaret.shmitchell@gmail.com
The Aether

Abstract

Transformer-based models have pushed state of the art in many areas of NLP, but our understanding of what is behind their success is still limited. This paper is the first survey of over 150 studies of the popular BERT model. We

Harvard Data Science Review • Issue 1.2, Fall 2019

Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From An Explainable AI Competition

Cynthia Rudin, Joanna Radin

Breaking BERT: Understanding its Vulnerabilities for Named Entity Recognition through Adversarial Attack

Anne Dirkson Suzan Verberne Wessel Kraaij
Leiden Institute of Advanced Computer Science, Leiden University
Niels Bohrweg 1, 2333 CA Leiden, the Netherlands
son, s.verberne, w.kraaij}@liacs.leidenuniv.nl

GPT-3 Understands Nothing

Fabio Tollon Follow
Feb 22 · 7 min read ★

Challenges in Detoxifying Language Models

nes Welbl* Amelia Glaese* Jonathan Uesato* Sumanth Dathathri*
John Mellor* Lisa Anne Hendricks Kirsty Anderson
Pushmeet Kohli Ben Coppin Po-Sen Huang*
DeepMind
{welbl,glamia,juesato,sdathath,johnme,posenhuang}@deepmind.com

Open Problems – iii

Sophisticated generative natural language processing (NLP) processing models such as GPT-3 also have a tendency to 'hallucinate' this kind of deceptive data. In part, this is because language models require the capability to rephrase and summarize long and often labyrinthine tracts of text, without any architectural constraint that's able to define, encapsulate and 'seal' events and facts so that they are protected from the process of semantic reconstruction.

Therefore the facts are not sacred to an NLP model; they can easily end up treated in the context of 'semantic Lego bricks', particularly where complex grammar or arcane source material makes it difficult to separate discrete entities from language structure

Human: Who wrote the novel Bleak House?
AI: Charles Dickens.

Who wrote the novel Bleak House?
AI: I don't remember.

Q: What did Albert Einstein say about dice?
A: "I never throw dice."

Q: What did Albert Einstein say about time?
A: "If a person sits for a minute he will be an hour behind."

Q: What did Albert Einstein say about his theory?
A: "It's not the theory that's difficult, it's the imagination."

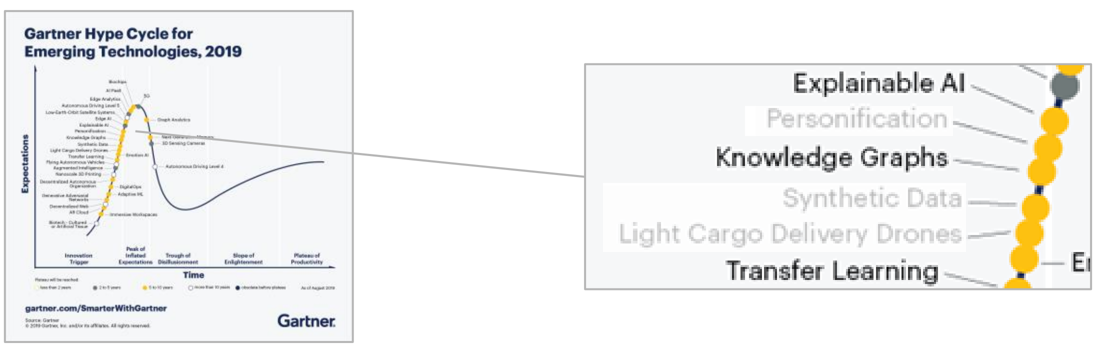
Davinci engine in GPT-3: fragility and false Einstein quotes

Article: Super Bowl 50
Paragraph: *"Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV."*
Question: *"What is the name of the quarterback who was 38 in Super Bowl XXXIII?"*
Original Prediction: John Elway
Prediction under adversary: Jeff Dean

Distracting phrases for machine reading comprehension systems

Towards Mixed AI

- Integration of neural and symbolic approaches
  – NLP solutions and Semantic Web technologies (SWTs) are increasingly combined to solve their mutual weaknesses and achieve better results
  – The joint use of subsymbolic and symbolic AI is expected to be central in many applications
  – Mixed AI aims to reconcile the stochastic nature of learning with the logical representation of extracted knowledge and reasoning
  – Explainability, knowledge injection, common sense, small data, transfer learning
- Knowledge graphs (KGs) are constantly spreading
  – WordNet, Freebase, DBpedia, YAGO are among the most widely used
  – KGs model large collections of semantically interconnected data
  – Represent knowledge, integrate it and perform inference based on automatic reasoning
  – One of the most promising technologies of the next decade ([2019 Gartner's Hype Cycle]())



A new paradigm

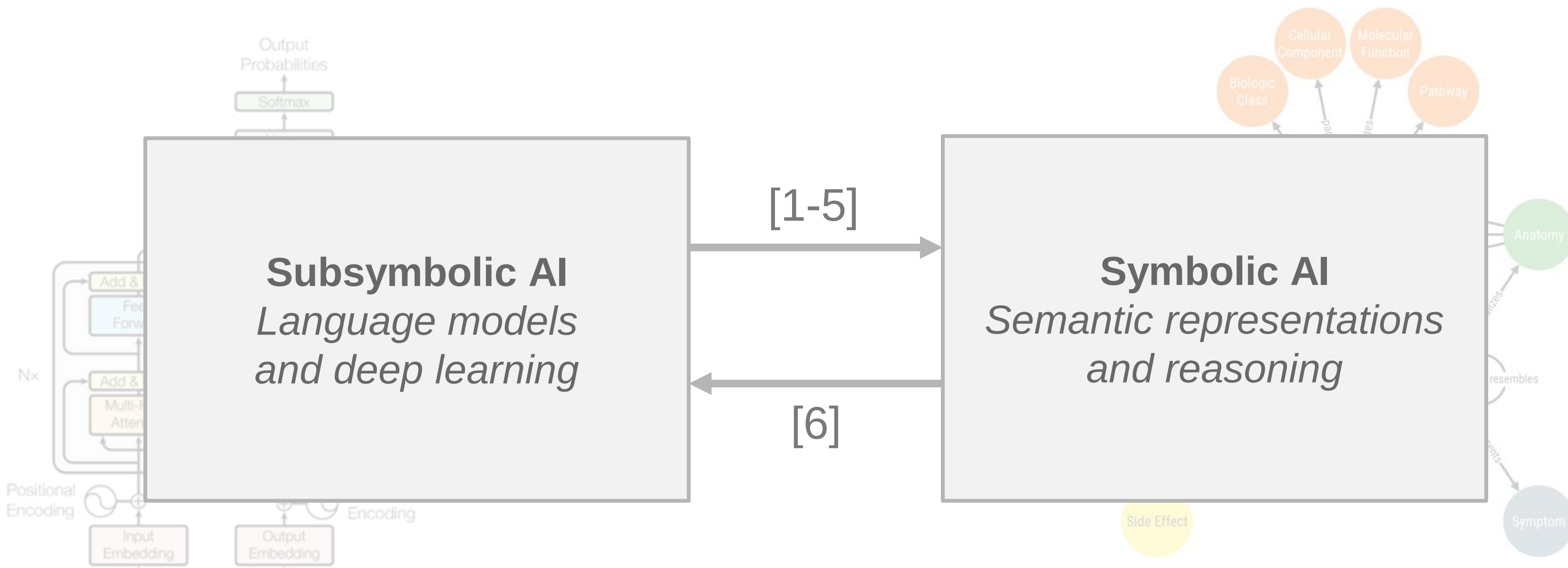
Just as biology does not choose between nature versus nurture, but uses them jointly to build wholes which are greater than the sums of their parts, we too should ***reject the notion that structure and flexibility are somehow at odds or incompatible****.*

- Peter Battaglia, DeepMind Researcher, 2018

Considering all topics that are currently at the focus of public attention (such as Artificial Intelligence, Machine Learning, Data Science, Natural Language Processing, Semantic Web), in the immediate future, I cannot see anything on top of that but rather ***a fusion of all of that****.*

- Andreas Blumauer, Semantic Web Company CEO, 2018

Research Goal and Activities



[1] Frisoni, G., Moro, G. and Carbonaro, A., 2020. Learning Interpretable and Statistically Significant Knowledge from Unlabeled Corpora of Social Text Messages: A Novel Methodology of Descriptive Text Mining. In *DATA* (pp. 121-132). **Best Paper Award**.

[2] Frisoni, G. and Moro, G., 2020, July. Phenomena Explanation from Text: Unsupervised Learning of Interpretable and Statistically Significant Knowledge. In *International Conference on Data Management Technologies and Applications* (pp. 293-318). Springer, Cham.

[3] Frisoni, G., Moro, G. and Carbonaro, A., 2020. Unsupervised Descriptive Text Mining for Knowledge Graph Learning. In *KDIR* (pp. 316-324).

[4] Frisoni, G., Moro, G. and Carbonaro, A., 2020, April. Towards Rare Disease Knowledge Graph Learning from Social Posts of Patients. In *The International Research & Innovation Forum* (pp. 577-589). Springer, Cham.

[5] Frisoni, G., Moro, G. and Carbonaro, A., 2021. A Survey on Event Extraction for Natural Language Understanding: Riding the Biomedical Literature Wave. *IEEE Access*, vol. 9, pp. 160721-160757.

[6] Frisoni, G., Moro, G., Carbonaro, A. and Carlassare G., 2021. Unsupervised Event Graph Similarity Learning on Biomedical Literature. *Sensors*, vol. 22 (1).

… Many others coming

POIROT:
Phenomena explanation from text

The research was developed by Gianluca Moro who is the author of the contribution that is also included in his text mining subject at the University of Bologna since the 2014/15 academic year and applied for the discovery and explanation of the reasons that contribute to cause aircraft accidents from the raw textual reports of accidents collected by the US National Transportation Safety Board (NTSB).

Phenomena explanation from text [1-2] – i

- Unsupervised methodology of descriptive text mining (called POIROT) to…
  - Learn semantic and statistically quantifiable correlations between relevant terms and documents (e.g., *"citrus fruit"* ↔ *"acid reflux": 87%*)
  - Construct phenomena descriptions through an original algorithm based on information retrieval and hypothesis testing, initially devised by professor Gianluca Moro
- Medical case study
  - Large online communities of patients aggregate to share experiences and to look for answers to questions in order to safely improve their health conditions
  - Medical literature typically works on samples and almost never considers the "voice of patients" expressed in the social sphere (i.e., their real-world)
  - Esophageal Achalasia (ORPHA:930), social posts analysis with PAO collaboration
- Tested on interpretable algebraic and probabilistic LMs
  - Latent Semantic Analysis (LSA), Probabilistic Latent Semantic Analysis (pLSA), and Latent Dirichlet Allocation (LDA) ← much lower embedding dimensions than neural counterparts
  - We automatically discovered scientific medical correlations from unlabeled patients' posts with ≈79% F1-score and recognized positive and negative aspects of Achalasia treatments

[1] Frisoni, G., Moro, G. and Carbonaro, A., 2020. Learning Interpretable and Statistically Significant Knowledge from Unlabeled Corpora of Social Text Messages: A Novel Methodology of Descriptive Text Mining. In *DATA* (pp. 121-132). **Best Paper Award.**

[2] Frisoni, G. and Moro, G., 2020, July. Phenomena Explanation from Text: Unsupervised Learning of Interpretable and Statistically Significant Knowledge. In *International Conference on Data Management Technologies and Applications* (pp. 293-318). Springer, Cham.

Phenomena explanation from text [1-2] – ii

What are the most difficult daily activities for patients?

Which foods cause or lighten a certain symptom?

Heller Myotomy or POEM?
1 Comments: 145

How long did it take you to get the diagnosis? It took me 2 years!!!
1 Comments: 218

Remedies for reflux? After 4 years from the surgery, I went back to being sick. What I do? The pain on the sternum is terrible :(
5 Comments: 49

From the moment of the surgery I am forced to take on the gastroprotector every morning. Do you also use it?
4 Comments: 112

Hi everyone, what is the best expert centre to perform the POEM intervention in your opinion?
4 Comments: 99

Today it took me two hours to complete the meal. I feel the food stuck in my throat. Does this happen to you too?
5 Comments: 88

POIROT

Esophageal Achalasia, AMAE Onlus Facebook Group
6,917 posts and 61,692 comments between 21/02/2009 and 05/08/2019 (10 years), ≈2000 users
Data downloaded with Facebook Graph API

A semantic space of posts and terms

Examples taken from our dataset

I would like to know if any of you have experienced the new endoscopic POEM technique performed at Gemelli University Hospital… I would like to know what is about, what are the benefits you have obtained

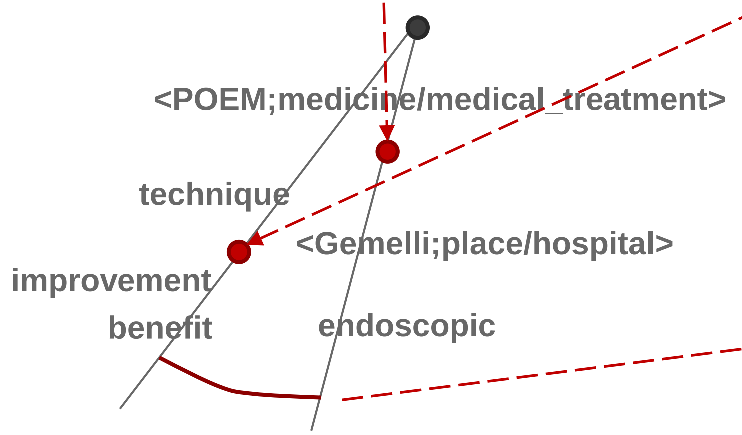
5 Comments: 99

Same meaning

Do you know POEM technique? They say it is possible to achieve significant improvements in terms of therapy… It is true?

5 Comments: 128

$$C_k = U_k \Sigma_k V_k^T$$



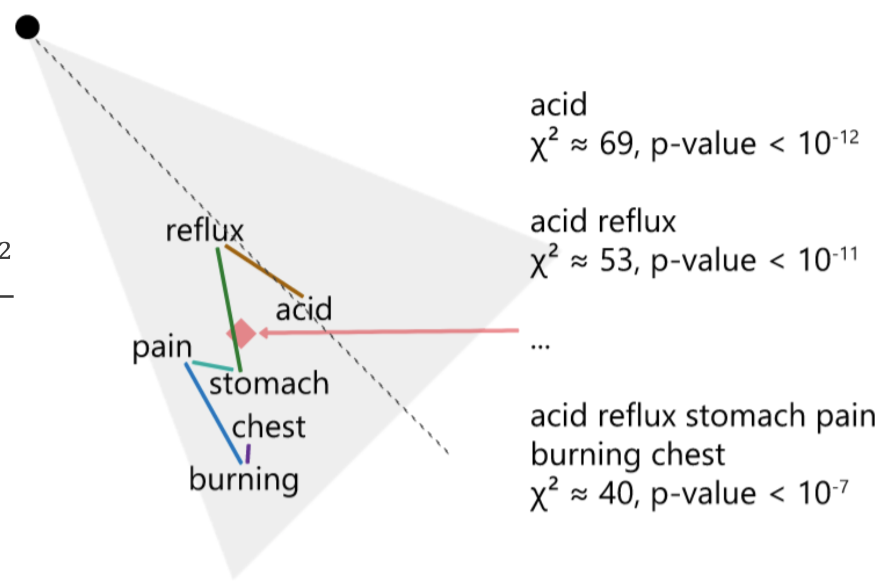
The two posts are recognized as semantically correlated (high cosine similarity)

Building a phenomena description

- At each step, the query is enriched with a semantically close term, and the correlation with the class (e.g., bad opinion) tends to decrease
- Fold-in (transform a query vector in a row of the matrix V_k)
- Formal verification of a correlation between the query *q* and the class *c*
  - Chi-squared (χ2) test in conjunction with R-precision (number of instances of class *c*)

$$q_k = q^T U_k \Sigma_k^{-1}$$

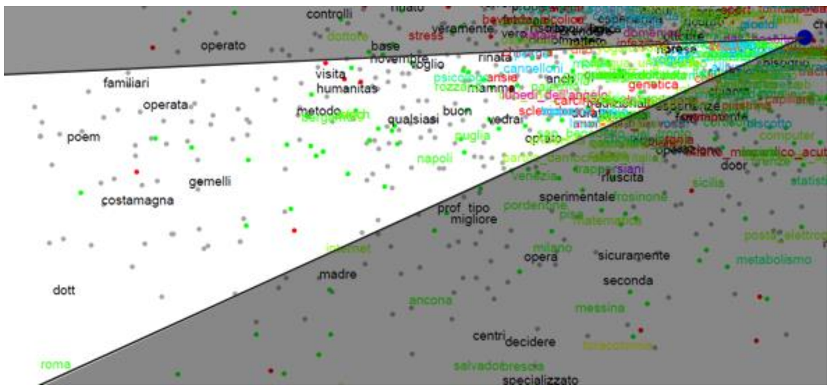
$$\chi^2(D, q, c) = \sum_{e_q \in \{0,1\}} \sum_{e_c \in \{0,1\}} \frac{(N_{e_q e_c} - E_{e_q e_c})^2}{E_{e_q e_c}}$$



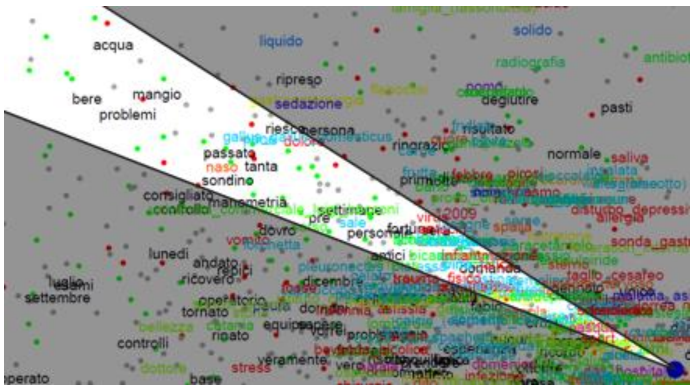
POIROT results

Detected with P-value < 0.01

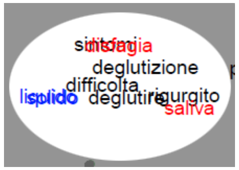
Treatment – Expert Centre - Doctors
poem operated gemelli dott costamagna familiari



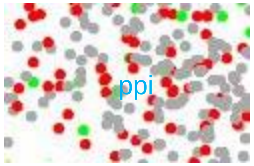
Symptom – Activity
Diagnostic Test – Medical Equipment
problems eat drink water
nose tube manometry examination



Absence of peristalsis → lack of cardias relaxation (neural dysfunctions) → dysphagia (both solid and liquid)



PPI ↔ Negative Opinion



	Pos Explanation	Neg Explanation
Heller-Dor	equipe, dr, costantini, salvador, padua, antireflux, plastic	problems, drink, eat
POEM	rome, prof, gemelli, costamagna, equipe, familiari	reflux, inflammation, problems, liquid, pain, ppi, antacids

Results validated with the collaboration of the Department of Drug Sciences, University of Pavia

Let POIROT speak

- Generate explanations of medical concept sets from patient social posts with Linear Transformer
  - Fine-tuning of T5X-Base (state-of-the-art LM for natural language generation)
  - Performer-based linear attention
  - Pre-training on Commongen (dataset for generating sentences conditioned by concept sets of 3-5 terms)

Input: lansoprazole ppi drug acid
Output: «i am taking **lansoprazole ppi** which is an anal gesic **drug** for oral **acid**»

Input: achalasia esophagus nerve malfunction
Output: «the **malfunction** of the **nerve** endings of the **esophagus** is typical of **achalasia**»

Input: alcohol risk esophagitis
Output: «the **risk** of **esophagitis** is increasing with the consumption of **alcohol**»

Input: padua esophagus disease
Output: «**padua** is the best center for **esophagus disease** in europe»

POIROT for KG learning [3-4] – i

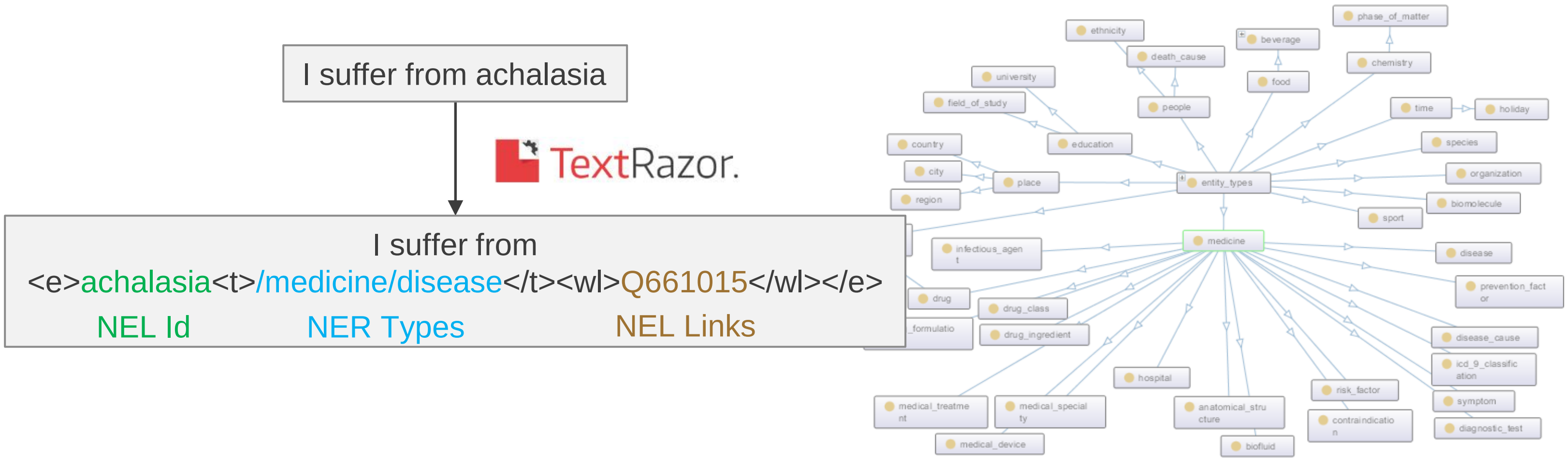
- Unfortunately, POIROT alone has some limitations…
- The general output is a flat list of correlated and quantified set of terms
  - E.g., <“dysphagia swallowing”: 86%; “alcohol acid reflux”: 81%>
- Terms are not semantically tagged and there are no links between different correlation sets → Not simple interpretation for users outside the domain
  - “poem” is a medical treatment ✘
  - “dysphagia” and “acid reflux” are both symptoms or diseases ✘
  - “acid reflux” and “GERD” are equivalent concepts in the context under consideration ✘
- If a user wants to investigate all the significant correlations between two or more types of entities, he is forced to check their instances individually
  - E.g., drug ↔ symptom: <“lansoprazole gerd”: ?>, <“aspirin headache”: ?”>, etc.

We showed that **POIROT** can be empowered for the **construction of KGs**, with resulting advantages in terms of expressive power, interpretability and interrogability

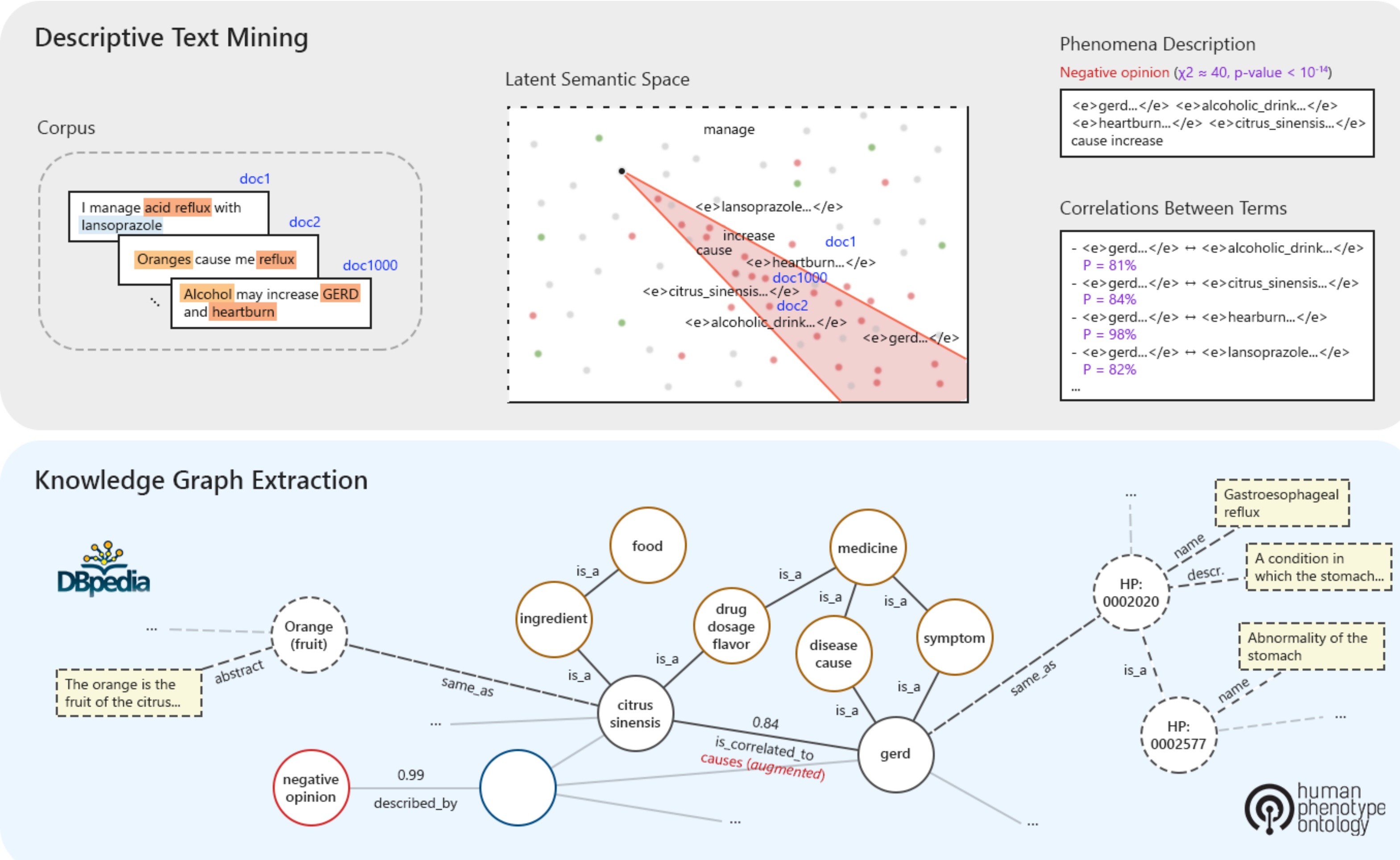
[3] Frisoni, G., Moro, G. and Carbonaro, A., 2020. Unsupervised Descriptive Text Mining for Knowledge Graph Learning. In *KDIR* (pp. 316-324).
[4] Frisoni, G., Moro, G. and Carbonaro, A., 2020, April. Towards Rare Disease Knowledge Graph Learning from Social Posts of Patients. In *The International Research & Innovation Forum* (pp. 577-589). Springer, Cham.

POIROT for KG learning [3-4] – ii

- A novel unsupervised and automatic technique of knowledge graph learning (KGL) from corpora of unstructured and unlabeled texts based on POIROT
  - Custom learning layer cake implementation w/ Protégé, Apache Jena, Fuzzy OWL 2, RServe
- We employed advanced named entity recognition (NER) and named entity linking (NEL) solutions in tandem with semantic web technologies
  - NER → fine-grained hierarchical taxonomies with 36,584,406 taxonomic types
  - NEL → disambiguation and linking to Wikipedia, DBpedia, Freebase, and PermID
  - We obtained data for >88,000 entities in posts and comments leveraging on TextRazor
  - Entity type mapping to a taxonomy for rare disease domains (reconciliate phase)



POIROT for KG learning [3-4] – iii



High-level process related to a medical domain as example, with Dbpedia and HPO as integrated knowledge bases

POIROT for KG learning [3-4] – iv

RAW UNLABELED POSTS/COMMENTS

POIROT IMPLEMENTATION

RServe

AUTOMATIC POPULATION SW

KG

protégé Open-source ontology editor

Jena Open-source Java framework for building Semantic Web applications

Fuzzy OWL 2 Fuzzy extension of OWL 2, based on fuzzy description logics (Protégé plug-in with extension based on annotation property)

RServe TCP / IP server that allows programs in various languages to connect to the currently open R session for submitting binary requests

POIROT for KG learning [3-4] – V

- Test the greater expressive power
  - SQWRL enables powerful queries on OWL resources, using a high-level abstract syntax for the representation of First Order Logic (FOL) and Horn-like rules
  - SQWRL queries + Pellet reasoner

Query 1
disease
↕
anatomical_structure

```
disease(?x) ^
anatomical_structure(?y) ^
correlation_has_subject(?c,?x) ^
correlation_has_object(?c,?y) ^
correlation_has_probability(?c,?P)
-> sqwrl:select(?x,?y,?P)
```

x	y	1 - p-value
esophagitis	esophagus	0.9947946
esophagitis	tissue (biology)	0.9839414
esophagitis	vocal cords	0.9833640
dysphagia	cardias	0.9775162
...		

Query 2
symptom
↕
"wine"

```
symptom(?x) ^
correlation_has_subject(?c,?x) ^
correlation_has_object(?c,wine) ^
correlation_has_probability(?c,?P)
-> sqwrl:select(?x,?P)
```

x	1 - p-value
gerd	0.9616342
nausea	0.9532385
...	

Query 3
drug
↕ pos
drug

```
drug(?x) ^
drug(?y) ^
correlation_has_subject(?c,?x) ^
correlation_has_object(?c,?y) ^
correlation_has_pos_probability(?c,?PP) ^
swrlb:greaterThan(?PP,0.9)
-> sqwrl:select(?x,?y,?PP)
```

x	y	1 - p-value
ppi	antacid	$1 - 2.041565e^{-114}$
antacid	riopan	$1 - 1.195385e^{-65}$
nifedipine	lansoprazole	$1 - 7.170313e^{-56}$
...		

- Correlations are interconnected
- NER labels for greater interpretability
  - E.g., "gemelli" is_a /medicine/hospital
- Meta-level queries with unbounded terms
  - E.g., drug ↔ symptom, drug ↔ "chest pain"
- Knowledge augmentation via linked open data

Events for Natural Language Understanding

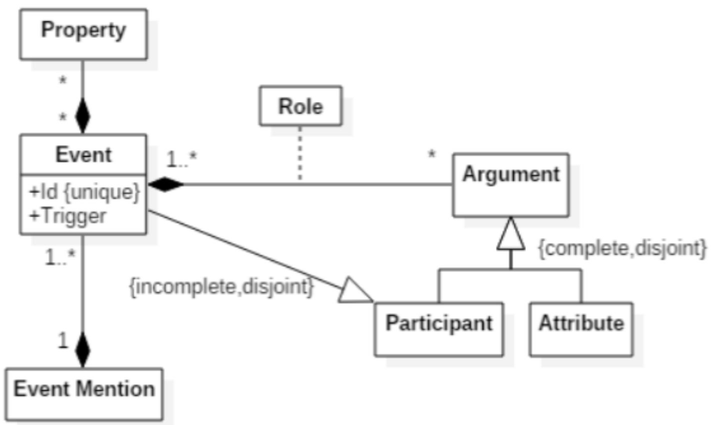
Events – i

- In the NLP field, information extraction is notably evolving from entities and pairwise relations to events
  - Events are an expressive method of capturing natural language statement semantics, enabling the formal representation of sophisticated processes

> *An event is a specific occurrence of something that happens and involves an arbitrary number of attributes and participants covering a specific semantic role, depending on the event type. The interaction (i.e., dynamic relation) modeled by an event represents or leads to some state change.*

- **Biomedical examples:** gene expression, protein binding, regulation, infection, cell death, adverse drug reactions
- **Non-biomedical examples:** terrorist attacks, transfer money, start-org

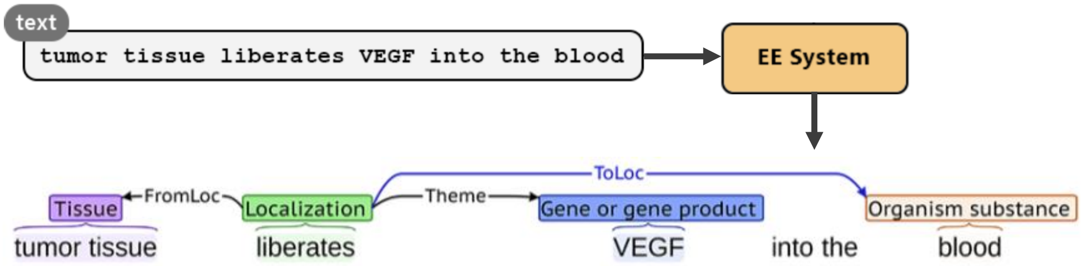
---



- **Event trigger:** textual mention that more clearly expresses event occurrence (e.g., “stimulates”, “regulates”)
- **Event argument:** entity mention, value, or another event that serves as a participant or attribute, fulfilling a specific role that characterizes its contribution (e.g., “NF-kappa”, “glucose”)
- **Argument role:** semantic relationship between an argument and an event, like “Cause” (what is responsible for event occurring) and “Theme” (what is affected)
- **Event modifier:** property that reshapes the described interaction (e.g., negation, speculation)
- **Event mention:** text span describing an event all its components

Events – ii

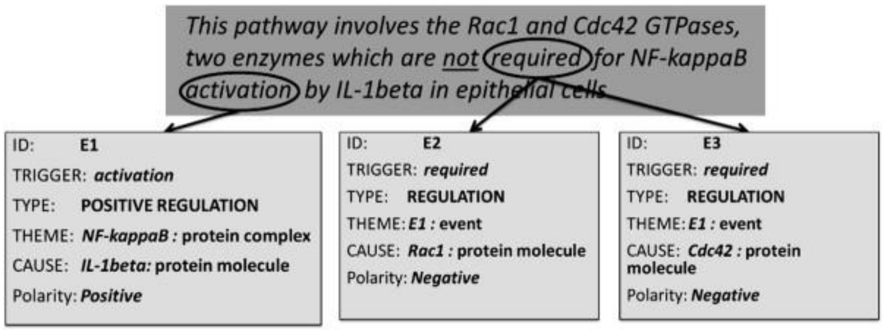
- Semantic parsing
  - Event extraction (EE) automatically creates **structured symbolic representations from large unstructured text corpora**
  - EE allows **deriving exhaustive, interpretable, concise, unambiguous and quantifiable interactions** otherwise buried in text spans and not readily available for further usage
- Events vs plain sentences
  - **Language is highly ambiguous**, with multiple ways to express the same concept unit, frequent high-level linguistic phenomena, untold background knowledge
  - The superficial organization of a sentence is almost irrelevant for the identification of its real and deeper semantic content, determined instead by **techtogrammatics** [7]
  - Events allow us to **remove noise** and focus only on **unambiguous relational knowledge involving relevant entities**



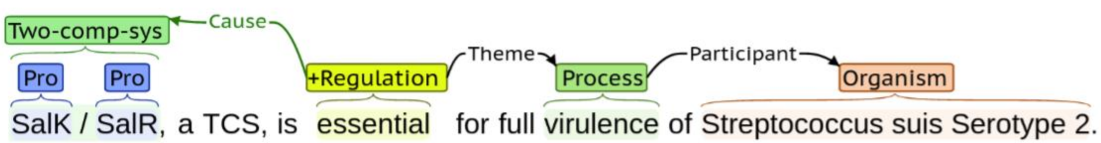
[7] Sgall, P., Hajicová, E., Hajicová, E., Panevová, J. and Panevova, J., 1986. The meaning of the sentence in its semantic and pragmatic aspects. Springer Science & Business Media.

Events – iii

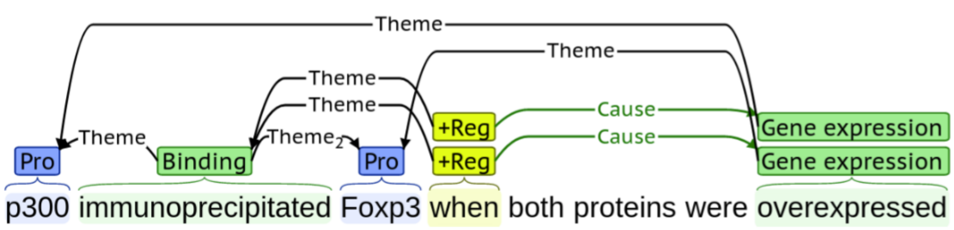
- Challenges
  - Directed links, N-ary relations, Multiple fine-grained types, Entity overlapping and coreference resolution, event double tagging, long-range dependencies, high-level linguistic phenomena



Event with negative polarity



Entity overlapping (simplified from PMC2358977)



Entity coreference resolution and double tagging (PMC1447668)

Events – iv

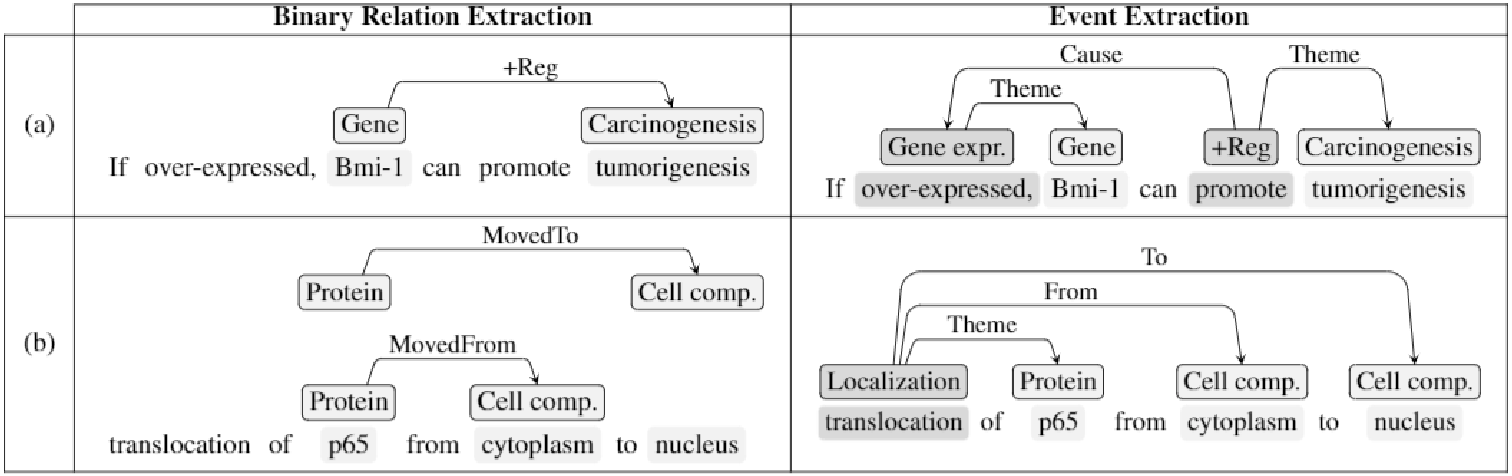
- Event schema
  - Closed-domain EE systems discover desired events from free-text, according to specified **event schemas** (i.e., target event types and participants constraints)
  - Each trigger, entity, and trigger-trigger/trigger-entity association is assigned to its type, according to a predefined **ontology**

Event Type	Arguments
Pathological	
Mutation	*Theme*(GGP), *AtLoc?*(Anatomical/Pathological), *Site?*
Metastasis	*Theme?*(Anatomical/Pathological), *ToLoc?*(Anatomical/Pathological)
Infection	*Theme?*(Anatomical/Pathological), *Participant?*(Organism)
Molecular	
Gene expression	*Theme+*(GGP)
Transcription	*Theme*(GGP)
General	
Binding	*Theme+*(Molecule), *Site?*(Protein or DNA domain/region)
Regulation	*Theme*(Any), *Cause?*(Any)
Positive Regulation	*Theme*(Any), *Cause?*(Any)
Negative Regulation	*Theme*(Any), *Cause?*(Any)
Planned process	*Theme**(Any), *Instrument**(Any)

*Example of event schema from **Cancer Genetics**. The gray labels are shown for organization purposes and are not included in the target types. The format **Arg(T)** indicates an event taking an argument **Arg** which should identify an entity of type **T**. The affixes **?**, ***** and **+** denote the cardinalities 0..1, 0..N, and 1..N, respectively.*

Events – v

- EE vs other IE tasks
  - EE involves the identification of a **trigger**, a set of **multiple arguments** with a **semantic role**, and a set of optional **modifiers** or properties which reshape the described interaction
  - EE handles complex relation with **n-ary** participants, **nested** and **overlapping** definitions, often including **named entity recognition (NER)**
  - Directed links, multiple fine-grained types, event coreferences, long-range dependencies, etc.

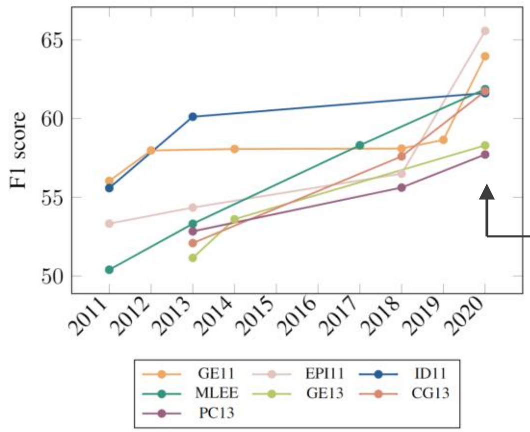


X Loses the condition for which the fact is correct

X Two binary incomplete facts instead of a single ternary fact

Events – vi

- Dataset limitations and importance
  - Event annotation process → **cost-prohibitive** (i.e., time-consuming, labor-intensive, domain knowledge demanding, and based on task-specific guidelines)
  - Resources are few, small, and precious for training
  - For instance, annotating the 9,372 sentences of the original Genia Event Corpus took **1.5 years** with five part-time annotators and two coordinators
- Solution taxonomy
  - Closed-domain vs open-domain, pipeline vs joint, sentence-level vs document-level, local features vs global ones, supervised vs unsupervised vs semi-supervised
- Performance



DeepEventMine (Trieu et al., 2020)

Event Applications – i

- Human languages frequently involves the description of real-world events
- Remarkably, EE can produce valuable structured information to facilitate a variety of downstream tasks and practical applications in diverse domains
- Biomedical EE help scientists to do research conveniently and provide inspiration and basis for the diagnosis, prevention, treatment, and new drug research
- Biomedical event models have strengthened…
  - Biomedical text mining
  - Knowledge base population and enrichment
  - Biomedical literature curation
  - Pathway curation
  - Biomedical literature-based knowledge discovery
  - Construction of biological interaction networks to narrow down the search space when exploring millions of biomedical publications
  - Diagnosis prediction based on sequences of clinical event sets
  - Machine reading comprehension over paragraphs describing biological processes, mapping questions to queries executed against predicted structured representations
  - …

Event Applications – ii

- Much more than biomedicine…
  - Single- and multi- document summarization
  - Evaluation metrics for natural language generation
  - Semantic search and information retrieval
  - Predicting people’s intents and reactions
  - Narrative comprehension for predicting what happens next in a story
  - Stock market predictions

RESEARCH-ARTICLE

Knowledge Graph-based Event Embedding Framework for Financial Quantitative Investments

Authors: Dawei Cheng, Fangzhou Yang, Xiaoyang Wang, Ying Zha

Authors Info & Affiliations

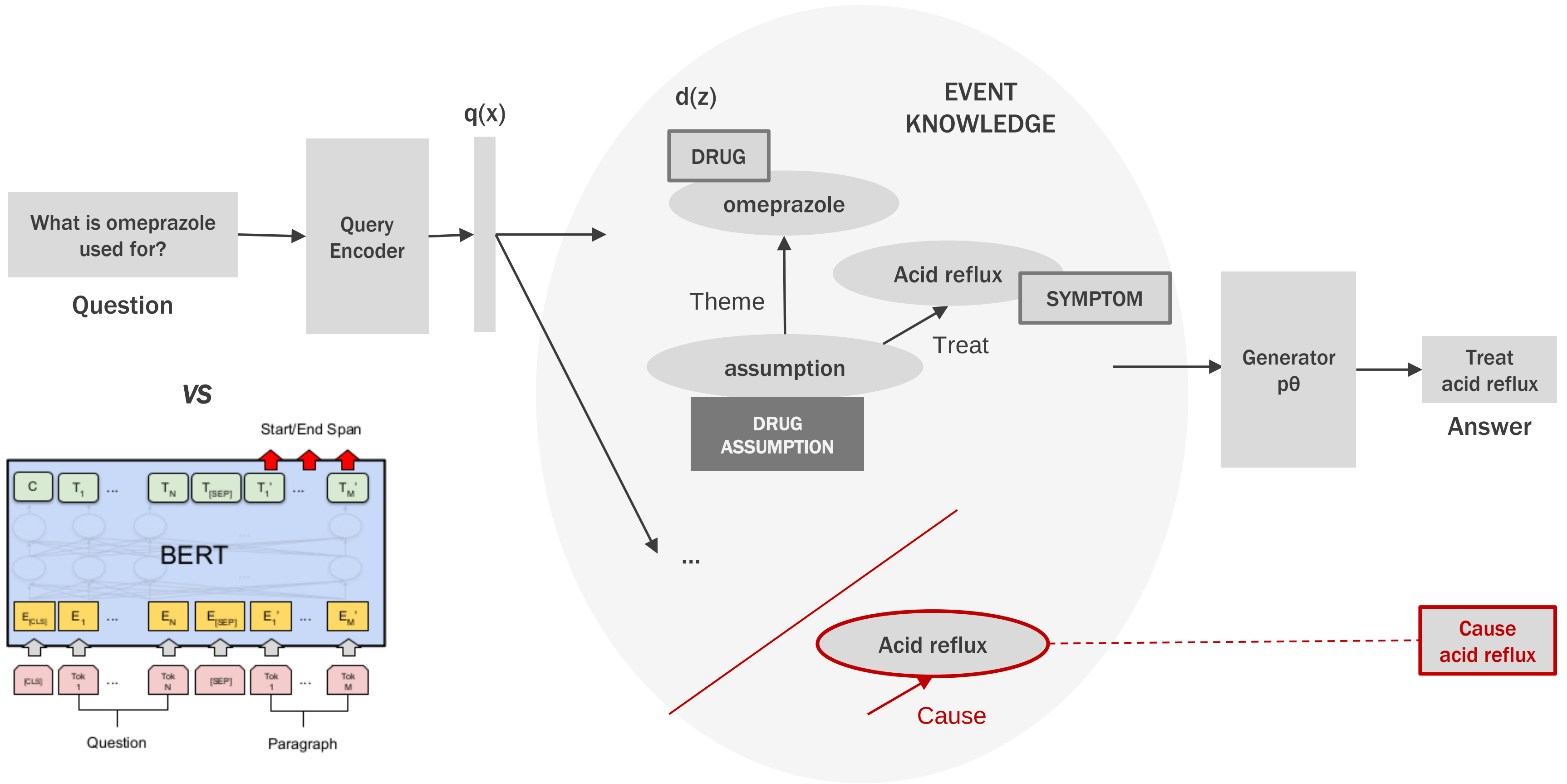
SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research
Retrieval • July 2020 • Pages 2221–2230 • https://doi.org/10.1145/3397271.34014

Deep Learning for Event-Driven Stock Prediction

Xiao Ding[†*], Yue Zhang[‡], Ting Liu[†], Junwen Duan[†]
[†]Research Center for Social Computing and Information Retrieval
Harbin Institute of Technology, China
{xding, tliu, jwduan}@ir.hit.edu.cn
[‡]Singapore University of Technology and Design
yue_zhang@sutd.edu.sg

Towards an event space

- Combining subsymbolic language models and symbolic events
  – Example inspired by RAG from Facebook [8]



[8] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.T., Rocktäschel, T. and Riedel, S., 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. arXiv preprint arXiv:2005.11401.

Events for Machine Reading Comprehension

Modeling Biological Processes for Reading Comprehension

Jonathan Berant*, Vivek Srikumar*, Pei-Chun Chen, Brad Huang and Christopher D. Manning
Stanford University, Stanford

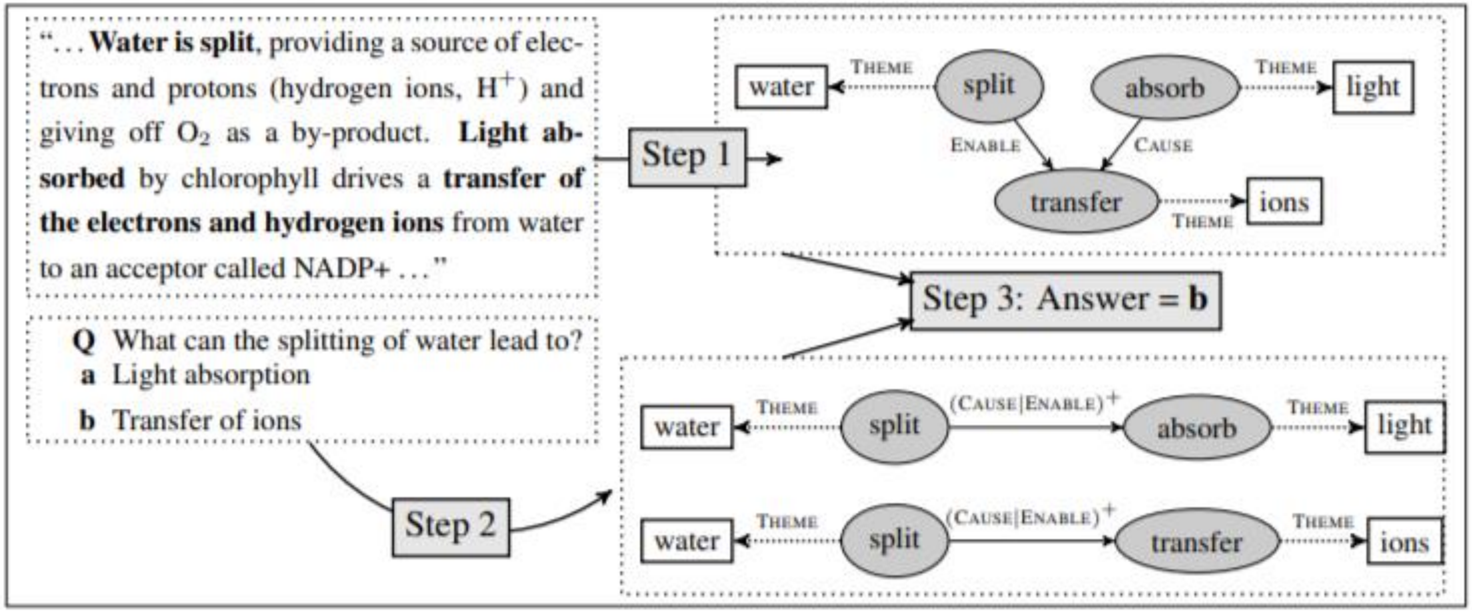


Figure 1: An overview of our reading comprehension system. First, we predict a structure from the input paragraph (the top right portion shows a partial structure skipping some arguments for brevity). Circles denote events, squares denote arguments, solid arrows represent event-event relations, and dashed arrows represent event-argument relations. Second, we map the question paired with each answer into a query that will be answered using the structure. The bottom right shows the query representation. Last, the two queries are executed against the structure, and a final answer is returned.

Our event-based contributions – i

The most comprehensive and up-to-date **survey on EE**, with 300+ reviewed works **[5]**

Unsupervised Event Graph Representation and Similarity Learning on Biomedical Literature [6]

- Cross-graph attention
- Whole-graph embeddings
- Leveraging both on structural and semantic properties

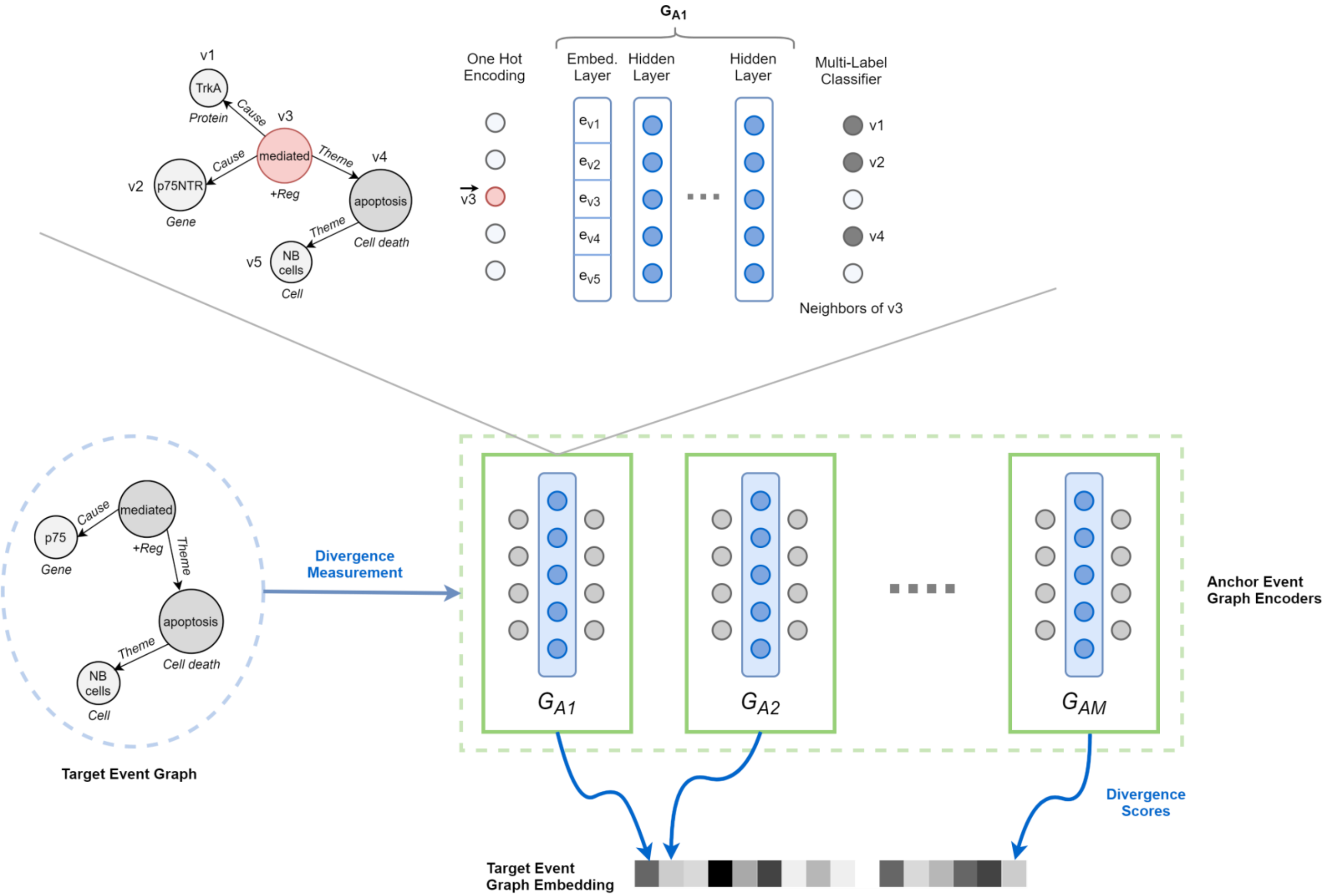
Event Graph Verbalization

- The first biomedical **event-text dataset** (obtained from 10 EE datasets), with more than 60K training instances
  - **Text-to-text format** based on a context free grammar (CFG)
- Baseline with state-of-the-art data-to-text NLG models, like T5 and BART
  - We obtained **63.09, 54.72, 60.12 (Rouge1/2/L)** with T5-Base, 100 epochs, batch size 16, linear attention
- **>2 Rouge F1 boost** (on average) by combining event-based tasks and single document summarization \w **multi-task learning**

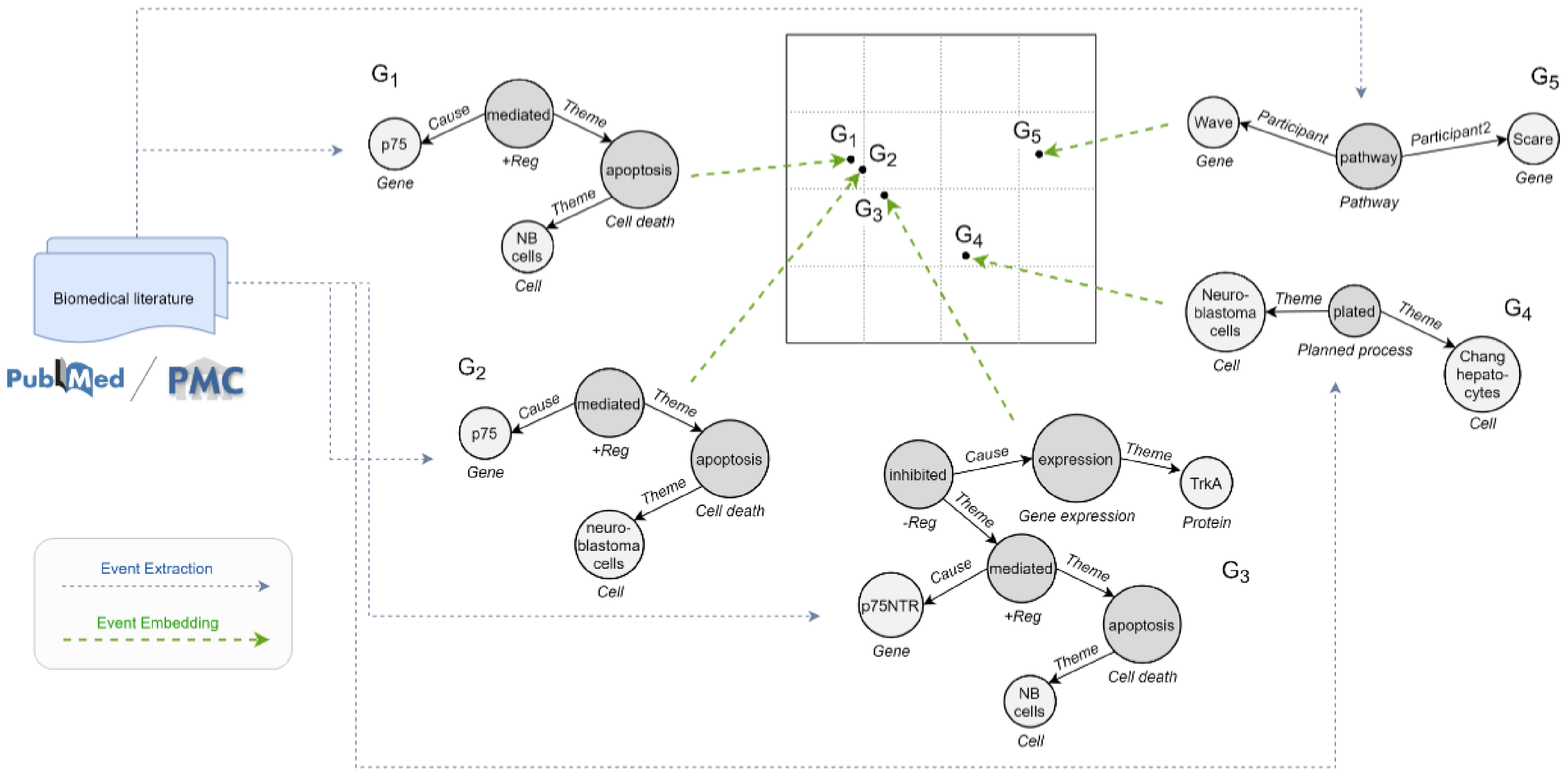
[5] Frisoni, G., Moro, G. and Carbonaro, A., 2021. A Survey on Event Extraction for Natural Language Understanding: Riding the Biomedical Literature Wave. Recently accepted as regular paper to *IEEE Access*.

[6] Frisoni, G., Moro, G., Carbonaro, A. and Carlassare G., 2021. Unsupervised Event Graph Representation and Similarity Learning on Biomedical Literature. Recently accepted as regular paper to *Sensors*.

Our event-based contributions – ii

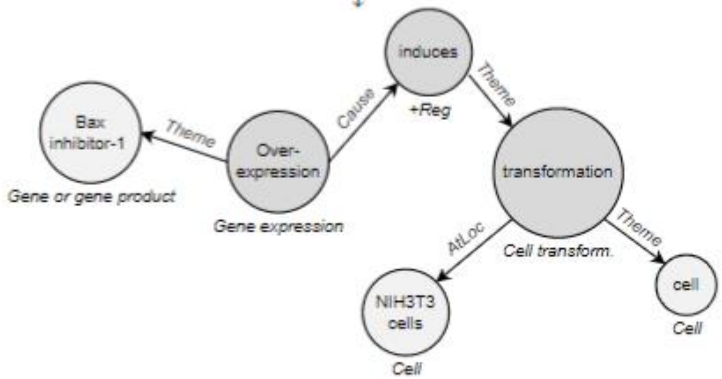
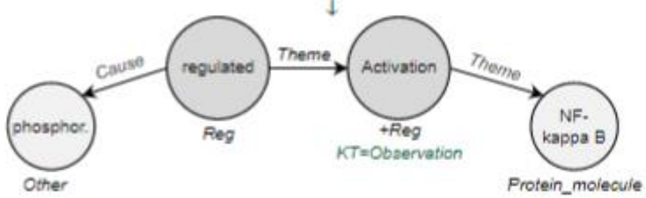


Our event-based contributions – iii



Highly related to graph neural networks (GNNs); hot topic

Our event-based contributions – iv

Event	Text
[Overexpression | Gene_expression] [Bax inhibitor-1 | Gene_or_gene_product | Theme = Overexpression] [[induces | Positive_regulation | Cause = Overexpression | Theme = transformation] [transformation | Cell_transformation] [NIH3T3 cells | Cell | AtLoc = transformation] [cell | Cell | Theme = transformation]] ↓ 	*Ground_truth* Overexpression of Bax inhibitor-1 (BI-1) induces cell transformation in NIH3T3 cells. *T5*[BioE2T] ✓ *BART*[BioE2T] ✓
[regulated | Regulation] [[Activation | Positive_regulation | KT = Observation | Theme = regulated] [NF-kappa B | Protein_molecule | Theme = Activation]] [phosphorylations | Other | Cause = regulated] ↓ 	*Ground_truth* Activation of NF-kappa B in vivo is regulated by multiple phosphorylations. *T5*[BioE2T] ✓ *BART*[BioE2T] Activation of NF-kappa B is regulated by phosphorylations and rapid degradation of its inhibitor I kappa B alpha.

Graph embedding

Edit the molecule to see how the prediction changes, or change the model params to load a different model. Select a different molecule in the scatter plot.

https://distill.pub/2021/gnn-intro/

Graphs and LMs

Relational graphs and language models are increasingly combined

Combining KGs and LMs – i

Barack's Wife Hillary:
Using Knowledge Graphs for Fact-Aware Language Modeling

Robert L. Logan IV* **Nelson F. Liu†§** **Matthew E. Peters§**
Matt Gardner§ **Sameer Singh***

[*Super Mario Land*] is a [*1989*] [*side-scrolling*] [*platform video game*] developed and published by [*Nintendo*] as a [*launch title*] for their [*Game Boy*] [*handheld game console*].

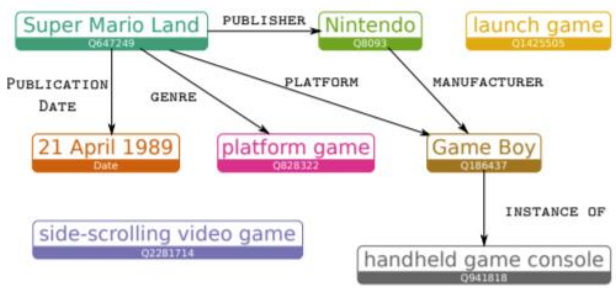


Figure 1: ***Linked WikiText-2* Example.** A localized knowledge graph containing facts that are (possibly) conveyed in the sentence above. The graph is built by iteratively linking each detected entity to Wikidata, then adding any relations to previously mentioned entities. Note that not all entities are connected, potentially due to missing relations in Wikidata.

- Knowledge graph language model (KGLM) for selecting and copying facts from a KG that are relevant to the context
- Recall facts encountered in text documents
- KGLM outperforms even very large language models in generating facts (i.e., completing sentences requiring factual knowledge)

	Input Sentence	**Gold**	**GPT-2**	**KGLM**
Both correct	Paris Hilton was born in ___	New York City	New	1981
	Arnold Schwarzenegger was born on ___	1947-07-30	July	30
KGLM correct	Bob Dylan was born in ___	Duluth	New	Duluth
	Barack Obama was born on ___	1961-08-04	January	August
	Ulysses is a book that was written by ___	James Joyce	a	James
GPTv2 correct	St. Louis is a city in the state of ___	Missouri	Missouri	Oldham
	Richard Nixon was born on ___	1913-01-09	January	20
	Kanye West is married to ___	Kim Kardashian	Kim	the
Both incorrect	The capital of India is ___	New Delhi	the	a
	Madonna is married to ___	Carlos Leon	a	Alex

Combining KGs and LMs – ii

Knowledge Graph-Augmented Abstractive Summarization with Semantic-Driven Cloze Reward

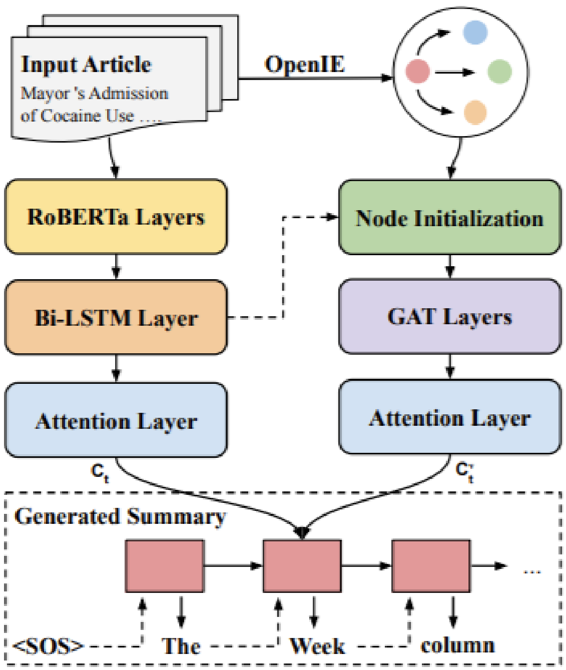


Figure 2: Our ASGARD framework with document-level graph encoding. Summary is generated by attending to both the graph and the input document.

- Text-Graph double channel encoding architecture, for a controlled generation with respect to both
- Abstractive single-document summarization with graph augmentation and semantic-driven reward (reinforcement learning)
- Structured representations allow the generation of more informative summaries with fewer unfaithful errors

Combining KGs and LMs – iii

KagNet: Knowledge-Aware Graph Networks for Commonsense Reasoning

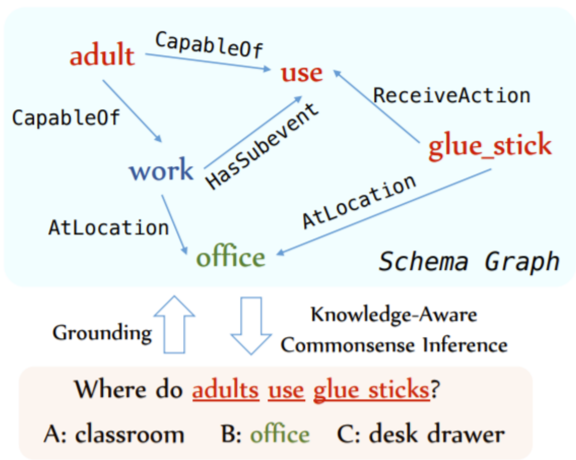
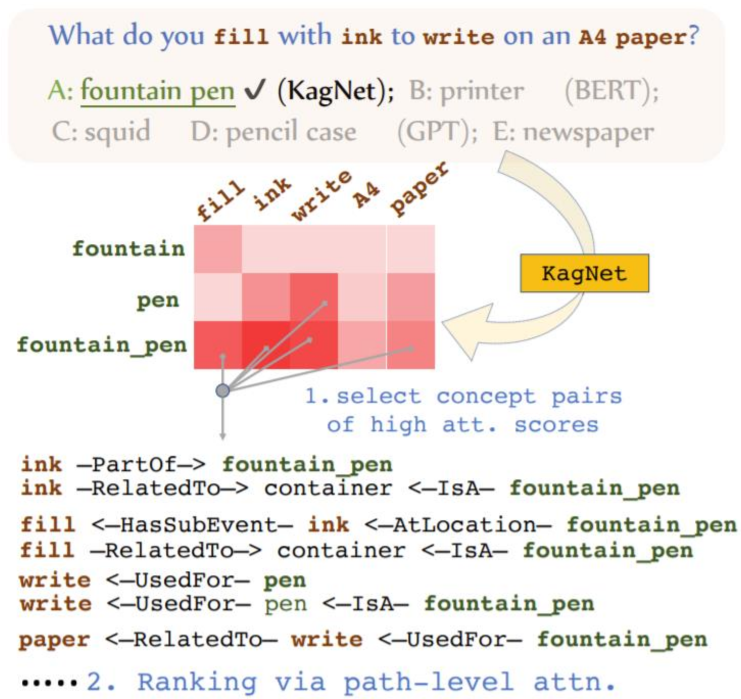


Figure 1: An example of using external commonsense knowledge (symbolic space) for inference in natural language commonsense questions (semantic space).



Question answering with external, structured commonsense KGs to boost performance and make explainable inferences

Combining KGs and LMs – iv

(COMET-)ATOMIC$^{20}_{20}$: On Symbolic and Neural Commonsense Knowledge Graphs

Jena D. Hwang[†*], Chandra Bhagavatula[†*], Ronan Le Bras[†], Jeff Da[†], Keisuke Sakaguchi[†], Antoine Bosselut[♠†] and Yejin Choi[♡†]

[†] Allen Institute for AI, WA, USA
[♡] Paul G. Allen School of Computer Science & Engineering, WA, USA
[♠] Stanford University, CA, USA

In this work, we posit that manually constructed CSKGs will never achieve the coverage necessary to be applicable in all situations encountered by NLP agents. Therefore, we propose a new evaluation framework for testing the utility of KGs based on how effectively implicit knowledge representations can be learned from them.

With this new goal, we propose ATOMIC$^{20}_{20}$, a new CSKG of general-purpose commonsense knowledge containing knowledge that is not readily available in pretrained language models. We evaluate its properties in comparison with other leading CSKGs, performing the first large-scale pairwise study of commonsense knowledge resources. Next, we show that ATOMIC$^{20}_{20}$ is better suited for training *knowledge models* that can generate accurate, representative knowledge for new, unseen entities and events. Finally, through human evaluation, we show that the few-shot performance of GPT-3 (175B parameters), while impressive, remains ~12 absolute points lower than a BART-based knowledge model trained on ATOMIC$^{20}_{20}$ despite using over 430x fewer parameters.

KG	Model	Accept	Reject	No Judgm.
ATOMIC$^{20}_{20}$	GPT2-XL	36.6	62.5	0.9
	GPT-3	73.0	24.6	2.5
	COMET(GPT2-XL)	72.5	26.6	0.9
	COMET(BART)	**84.5**	**13.8**	1.7
ATOMIC	GPT2-XL	38.3	61.2	0.4
	COMET(GPT2-XL)	64.1	34.7	1.2
	COMET(BART)	**83.1**	**15.3**	1.6
CONCEPTNET	GPT2-XL	50.3	42.1	7.7
	COMET(GPT2-XL)	74.5	19.0	6.4
	COMET(BART)	**75.5**	**17.9**	6.6
TRANSOMCS	GPT2-XL	**28.7**	**53.5**	17.8
	COMET(GPT2-XL)	26.9	60.9	12.2
	COMET(BART)	23.8	65.9	10.3

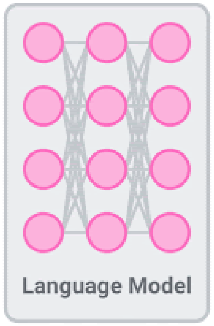
Table 6: Human evaluation of generation accuracy (%). Each model uses greedy decoding to generate the *tail* of 5K randomly-sampled test prefixes (*head*, *relation*) from each knowledge graph. GPT2-XL, GPT-3 and BART have 1.5B, 175B and 440M parameters, respectively.

- Fine-tuning LMs on KGs lead to a significant knowledge completion boost than GPT-3, with hundreds fewer parameters

Combining KGs and LMs – v

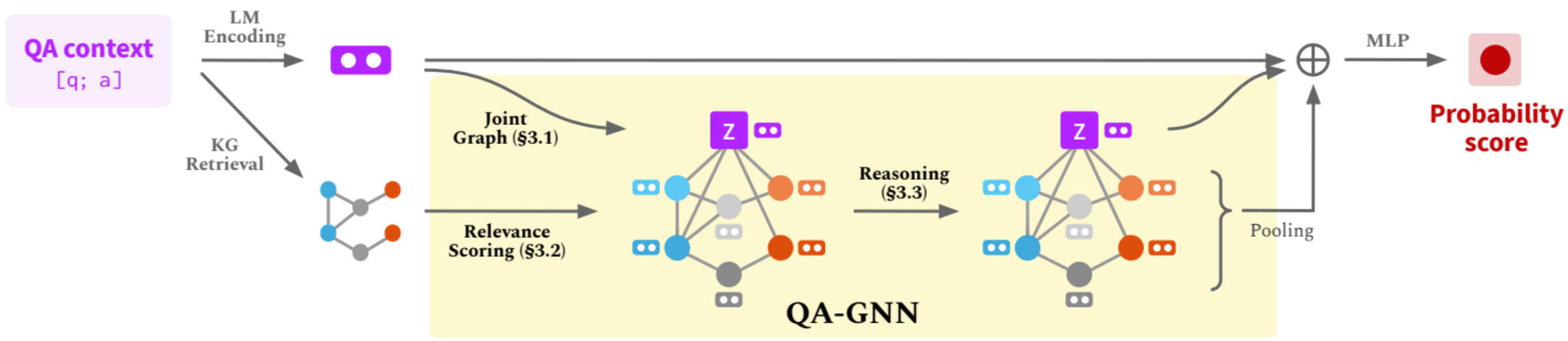
Knowledge Graph Based Synthetic Corpus Generation for Knowledge-Enhanced Language Model Pre-training

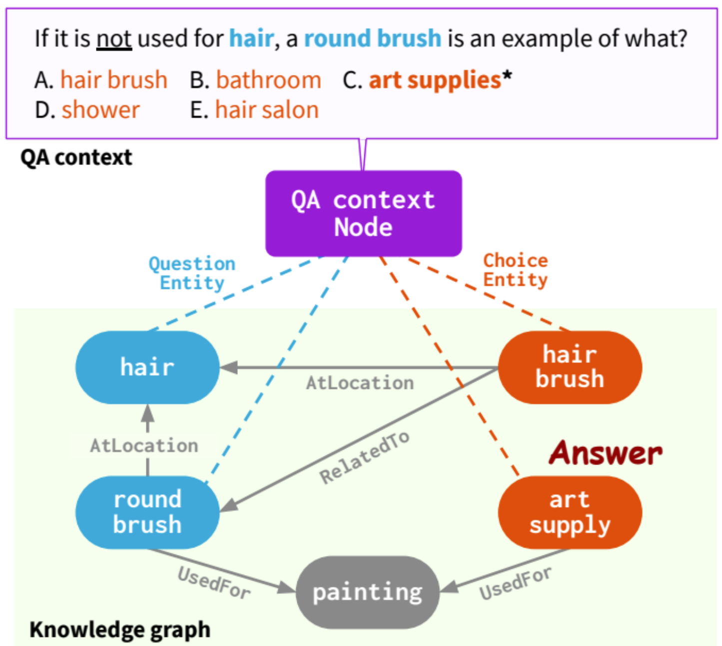
Oshin Agarwal[*1] **Heming Ge**[2] **Siamak Shakeri**[2] **Rami Al-Rfou**[2]
[1] University of Pennsylvania [2]Google Research
oagarwal@seas.upenn.edu, {hemingge, siamaks, rmyeid}@google.com



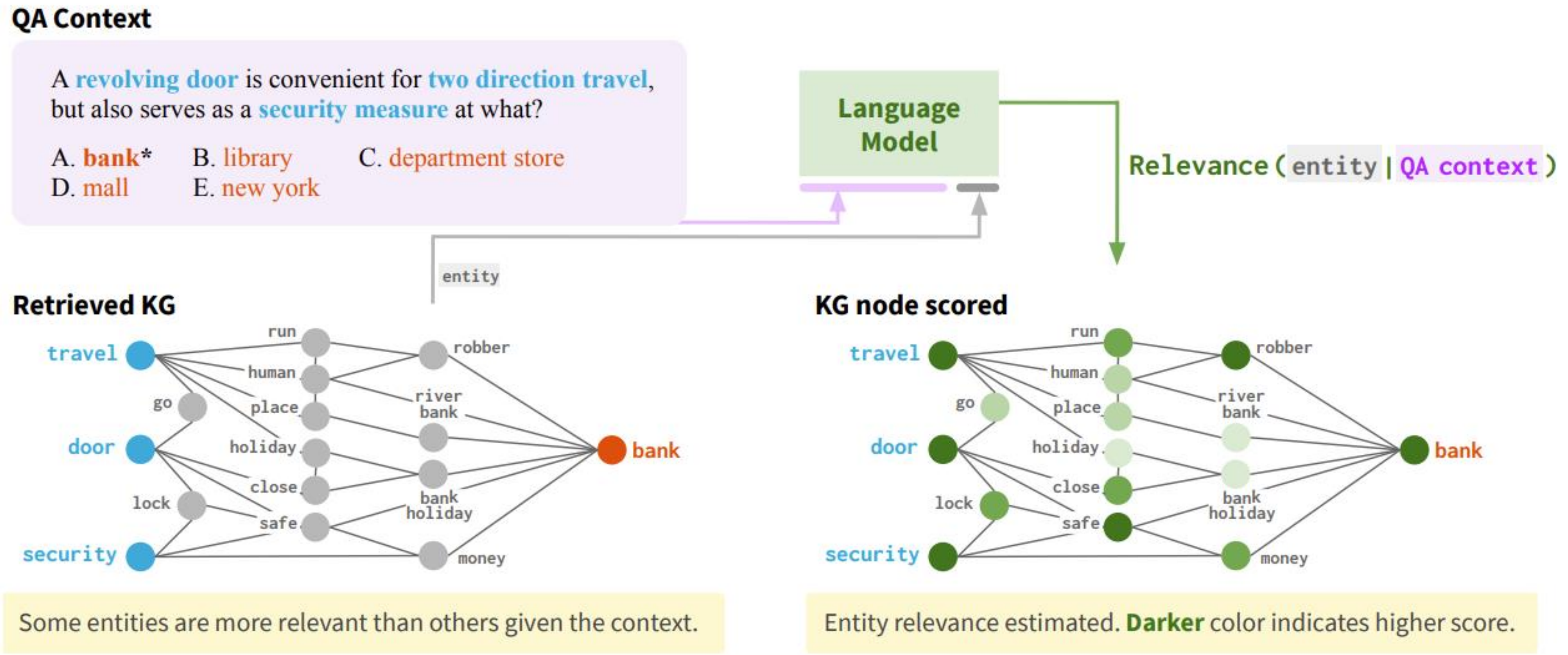
- KG verbalization to generate artificial documents for pre-training dataset extension
- New state-of-the-art in open question answering

QA-GNN – i


QA context
[q; a]
LM Encoding
KG Retrieval
Joint Graph (§3.1)
Relevance Scoring (§3.2)
Reasoning (§3.3)
Z
Pooling
MLP
QA-GNN
Probability score

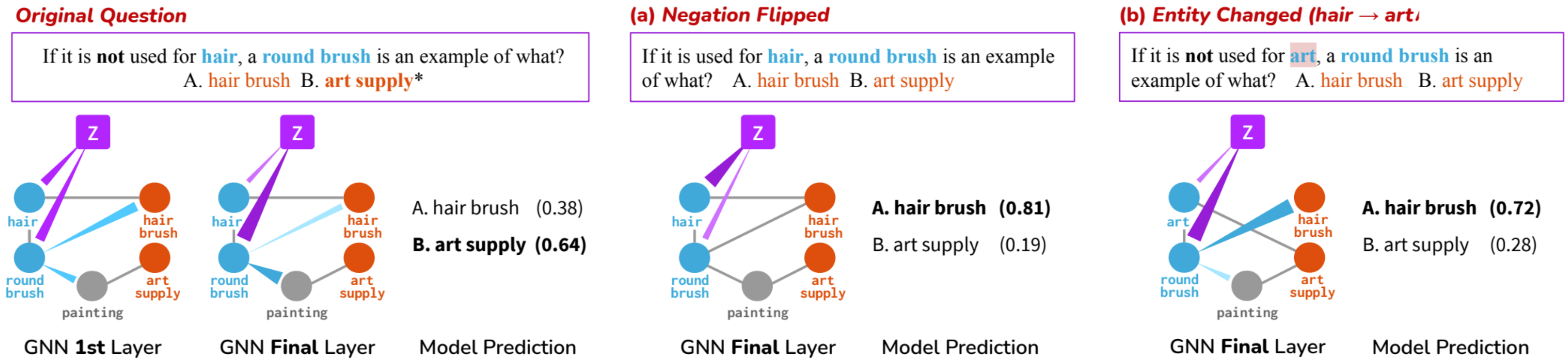

If it is not used for hair, a round brush is an example of what?
A. hair brush B. bathroom C. art supplies*
D. shower E. hair salon
QA context
QA context Node
Question Entity
Choice Entity
hair
hair brush
AtLocation
AtLocation
RelatedTo
round brush
Answer
art supply
UsedFor
painting
UsedFor
Knowledge graph

QA-GNN – ii



QA Context
A revolving door is convenient for two direction travel, but also serves as a security measure at what?
A. bank* B. library C. department store
D. mall E. new york
Language Model
Relevance (entity | QA context)
entity
Retrieved KG
travel
door
security
go
lock
run
human
place
holiday
close
safe
robber
river bank
bank holiday
money
bank
Some entities are more relevant than others given the context.
KG node scored
travel
door
security
go
lock
run
human
place
holiday
close
safe
robber
river bank
bank holiday
money
bank
Entity relevance estimated. Darker color indicates higher score.

QA-GNN – iii



Methods	Test
BERT-base (Devlin et al., 2019)	34.3
BioBERT-base (Lee et al., 2020)	34.1
RoBERTa-large (Liu et al., 2019)	35.0
BioBERT-large (Lee et al., 2020)	36.7
SapBERT (Liu et al., 2020a)	37.2
SapBERT + QA-GNN (**Ours**)	**38.0**

Table 6: **Test accuracy on *MedQA-USMLE*.**

Some possible thesis

What’s next?

Thesis – i

- **Event aggregation**
  - Detect and aggregate semantically similar events
  - By doing so, for example, a doctor could instantly have access to the number of times patients have complained of a particular symptom or a drug side effect
- **Event-driven multi-document summarization**
  - Build an end-to-end system to summarize multiple input documents by performing event extraction, structural- and semantic-based event aggregation, and event verbalization
- **Integration of event knowledge withing LMs to boost several NLP tasks**
  - Event embedding (with or without graph neural networks) as a semantic layer to bolster factuality, interpretability, reduce toxicity, and hallucinations (e.g., question answering)
  - To what extent and in what way can event-driven modeling be superior to traditional language modeling?
- **Neural networks with event-based memories**
  - A big step forward on the road to high-performance, explainability, and knowledge injection could originate from incorporating human-like memories into neural networks
  - Provide the models with the ability to perform *Read* and *Write* operations
  - Node/edge editing of concept unit graphs like events could expose the knowledge learned by a model in a human-comprehensible way

Thesis – ii

- **You will learn to…**
  - Work with recent and promising deep learning solutions, like graph neural networks, neural networks with memories, and transformers
  - Master state-of-the-art LMs written in PyTorch, TensorFlow or Flax (also within open NLP communities like HuggingFace)
  - Define, train, and evaluate your own model on multiple GPUs, with Docker and Linux environments
  - Tackle research projects
  - Combine notions of different research areas (e.g., neural models and semantic web technologies) to build original solutions

While technology is important,
it’s what we do with it that truly matters.
- Muhammad Yunus,
Nobel Peace Prize Winner and microfinance pioneer

Thanks for the attention

(is all you need)

Notebook Colab

https://colab.research.google.com/drive/1IAQ7XVGRrpNHgM1YB_ZEs8vL07Z2En0P?usp=sharing