LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

IBM Research AI

The State of the Art Transformer Language Models on Knowledge Graph Construction from Text

Named Entity Recognition and Relation Extraction

Presented by: Jennifer D’Souza and Nandana Mihindukulasooriya

Leibniz Gemeinschaft

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

IBM Research AI

The State of the Art Transformer Language Models on Knowledge Graph Construction from Text

Named Entity Recognition and Relation Extraction

Presented by: Jennifer D’Souza and Nandana Mihindukulasooriya

Leibniz Gemeinschaft

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

IBM Research AI

The State of the Art Transformer Language Models on Knowledge Graph Construction from Text

Named Entity Recognition and Relation Extraction

Presented by: Jennifer D'Souza and Nandana Mihindukulasooriya

Leibniz Gemeinschaft

About us

Jennifer D’Souza

Technische Informationsbibliothek (TIB), Welfengarten 1B // 30167 Hannover
Open Research Knowledge Graph Team (https://www.orkg.org/orkg/)

Nandana Mihindukulasooriya

IBM Research AI
Knowledge Induction Team (https://research.ibm.com/artificial-intelligence)

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

IBM Research AI

The State of the Art Transformer Language Models on Knowledge Graph Construction from Text

Part 1: Named Entity Recognition (NER) Perspective

Presented by: Jennifer D'Souza, Postdoc at TIB Hannover
http://orkg.org | @orkg_org
Technische Informationsbibliothek (TIB)
Welfengarten 1B // 30167 Hannover

Leibniz Gemeinschaft

Named Entity Recognition

Named Entity Recognition (NER) is the process of finding entities (people, cities, organizations, dates, ...) in a text.

Named Entity Recognition

Named Entity Recognition (NER) is the process of finding entities (people, cities, organizations, dates, ...) in a text.

- find and classify names in text

Elvis Presley was born in 1935 in East Tupelo, Mississippi.

Named Entity Recognition

Named Entity Recognition (NER) is the process of finding entities (people, cities, organizations, dates, ...) in a text.

- find and classify names in text

Elvis Presley was born in 1935 in East Tupelo, Mississippi.

Named Entity Recognition

Named Entity Recognition (NER) is the process of finding entities (people, cities, organizations, dates, ...) in a text.

- find and classify names in text

Elvis Presley was born in $\underbrace{\text{1935}}_{\text{Date or Time expression}}$ in East Tupelo, Mississippi.

Person

Date or Time expression

Location

Named Entity Recognition

Named Entity Recognition (NER) is the process of finding entities (people, cities, organizations, dates, ...) in a text.

- find and classify names in text

Elvis Presley was born in 1935 in East Tupelo, Mississippi.

Person

Date or Time expression

City

State

Location

Named Entity Recognition

- is challenging because…
  - Variation of NEs – e.g. John Smith, Mr Smith, John
  - Ambiguity of NE types
    - John Smith (company vs. person)
    - May (person vs. month)
    - Washington (person vs. location)
    - 1945 (date vs. time)

Plan for Part I of II of the Talk

- Corpora
- Deep Learning Approaches
- Evaluations and State-of-the-Art

Plan for Part I of II of the Talk

- Corpora
- Deep Learning Approaches
- Evaluations and State-of-the-Art

IBM Research AI

Corpora

MUC

1995,
1998

Message Understanding Conference (MUC) Series

- organized in the 90s and financed by DARPA (Defense Advanced Research Projects Agency) to encourage the development of new and better methods of information extraction (IE).
- In this competition, many concurrent research teams competed against one another—required the development of standards for evaluation, e.g. the adoption of metrics like precision and recall, the introduction of NER and coreference resolution as automatic IE tasks.

IBM Research AI

Corpora

MUC

1995,
1998

MUC NER Specifications

- **Entity Types**. **Percent** and **Money** for NUMEX Tag; **Time** and **Date** Type for TIMEX Tag; and **Person**, **Location**, and **Organization** Types ENAMEX Tag

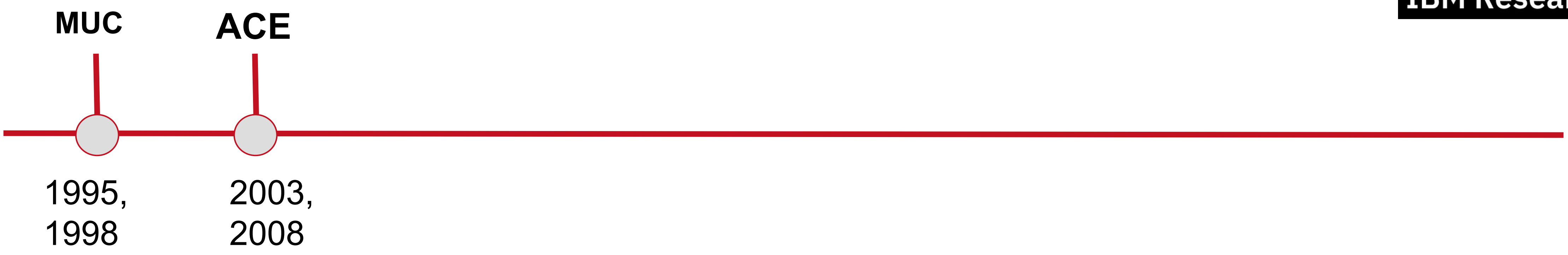
- **Genres**. Newswire

Reference

Grishman, Ralph, and Beth M. Sundheim. "Message understanding conference-6: A brief history." *COLING 1996 Volume 1: The 16th International Conference on Computational Linguistics*. 1996.

Corpora

IBM Research AI



Automatic Content Extraction (ACE) Series

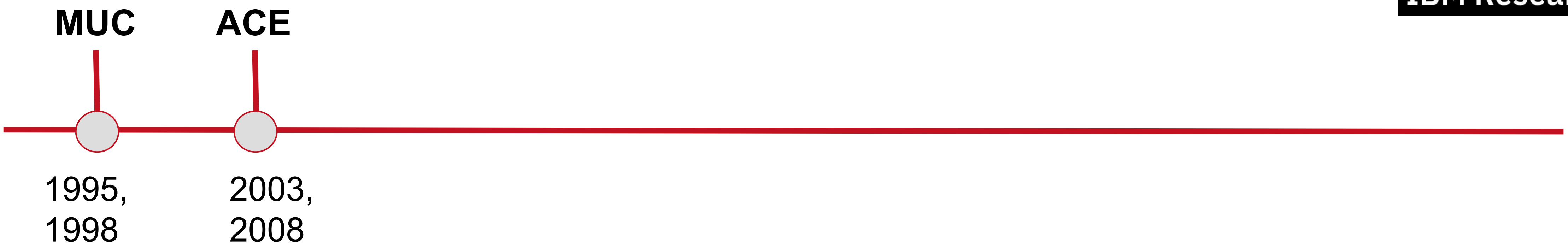
- a research program for developing advanced information extraction technologies convened by the NIST from 1999 to 2008, succeeding MUC
- There are three primary ACE annotation tasks corresponding to the three research tasks: Entity Detection and Tracking (EDT), Relation Detection and Characterization (RDC), and Event Detection and Characterization (EDC)

References

Doddington, George R., et al. "The automatic content extraction (ace) program-tasks, data, and evaluation." *Lrec*. Vol. 2. No. 1. 2004.

Corpora

IBM Research AI



ACE NER Specifications

- **Entity Types**. **Person**, **Organization**, **Location**, **Facility**, **Weapon**, **Vehicle**, **and Geo-political entities**

IBM Research AI

Corpora

MUC — 1995, 1998

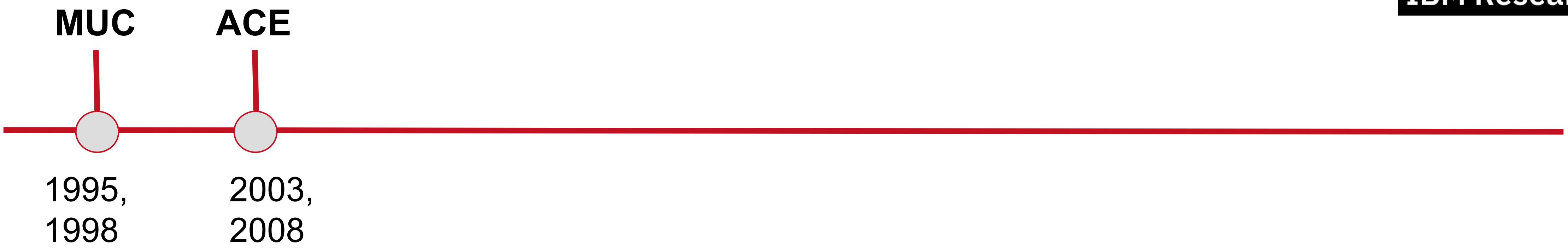
ACE — 2003, 2008

ACE NER Specifications

- **Entity Types**. **Person**, **Organization**, **Location**, **<u>Facility</u>**, **<u>Weapon</u>**, **<u>Vehicle</u>**, **and** **<u>Geo-political entities</u>**

IBM Research AI

Corpora

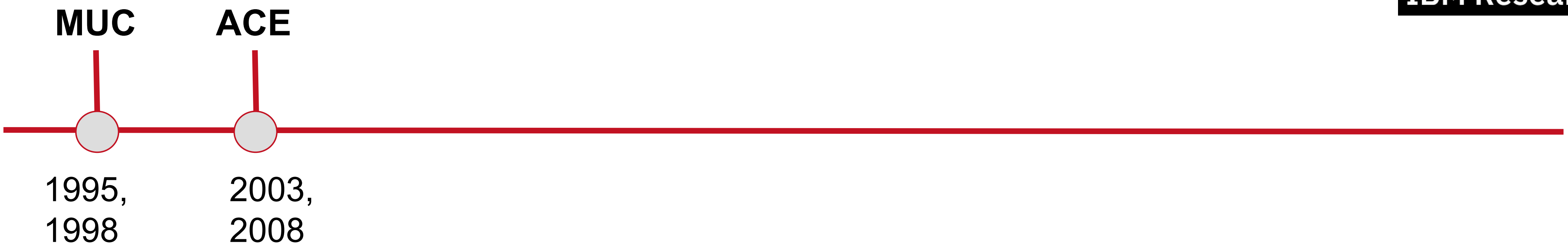


ACE NER Specifications

- **Entity Types**. **Person**, **Organization**, **Location**, **Facility**, **Weapon**, **Vehicle**, **and** **Geo-political entities**
- **Genres.** newswire, weblogs, Usenet newsgroups and bulletin boards, and transcripts of broadcast news, talk shows and conversational telephone speech

Corpora

IBM Research AI



ACE NER Specifications

- **Entity Types**. **Person**, **Organization**, **Location**, **Facility**, **Weapon**, **Vehicle**, **and** **Geo-political entities**
- **Genres.** newswire, weblogs, Usenet newsgroups and bulletin boards, and transcripts of broadcast news, talk shows and conversational telephone speech
- **Languages**. Arabic, English, and Chinese

Corpora

IBM Research AI

MUC 1995, 1998

ACE 2003, 2008

ACE NER Specifications

- **Entity Types**. **Person**, **Organization**, **Location**, **Facility**, **Weapon**, **Vehicle**, **and** **Geo-political entities**
- **Genres.** newswire, weblogs, Usenet newsgroups and bulletin boards, and transcripts of broadcast news, talk shows and conversational telephone speech
- **Languages**. Arabic, English, and Chinese
- **More information**. https://www.ldc.upenn.edu/collaborations/past-projects/ace/annotation-tasks-and-specifications

IBM Research AI

Corpora

Computational Natural Language Learning (CoNLL) Shared Task Series

- series begun in 2002 which overlaps in its timeline with ACE, nonetheless puts light on the change in focus on the community toward ***language-independent named entity recognition***

Reference

Sang, Erik F., and Fien De Meulder. "Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition." *arXiv preprint cs/0306050* (2003).

IBM Research AI

Corpora

CoNLL NER Specifications

- **Entity Types**. **Person**, **Organization**, **Location**, and **Miscellaneous**

Corpora

IBM Research AI

CoNLL NER Specifications

- **Entity Types**. **Person**, **Organization**, **Location**, **and Miscellaneous**
- **Genre.** Newswire

IBM Research AI

Corpora

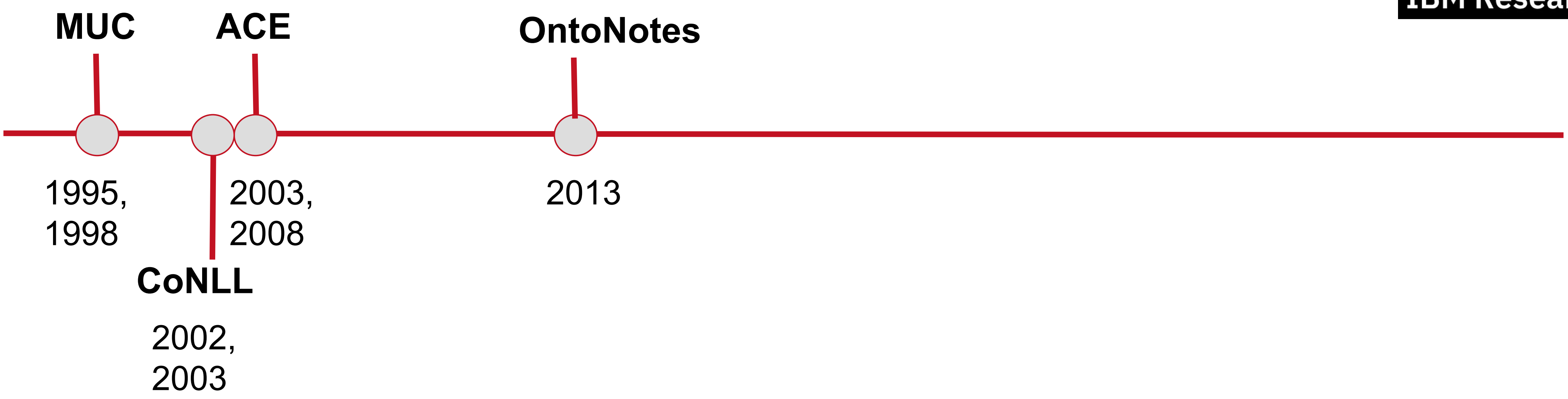
CoNLL NER Specifications

- **Entity Types**. **Person**, **Organization**, **Location**, and **Miscellaneous**
- **Genre.** Newswire
- **Languages**. Dutch and Spanish in CoNLL 2002; German and English in CoNLL 2003

TIB

IBM Research AI

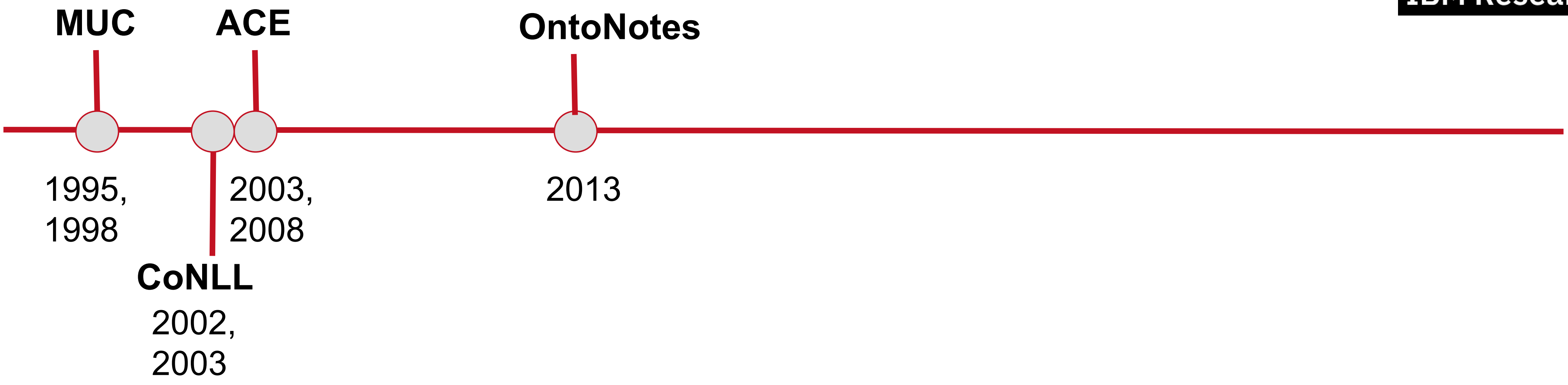
Corpora



OntoNotes Corpus ver. 5.0 (most popular corpora for NER)

- OntoNotes Release 5.0 is a collaborative effort between BBN Technologies, the University of Colorado, the University of Pennsylvania and the University of Southern California's Information Sciences Institute. The goal of the overall project was to annotate a large corpus comprising various genres of text (news, conversational telephone speech, weblogs, usenet newsgroups, broadcast, talk shows) in three languages (English, Chinese, and Arabic) with structural information (syntax and predicate argument structure) and shallow semantics (word sense linked to an ontology and coreference).

Corpora

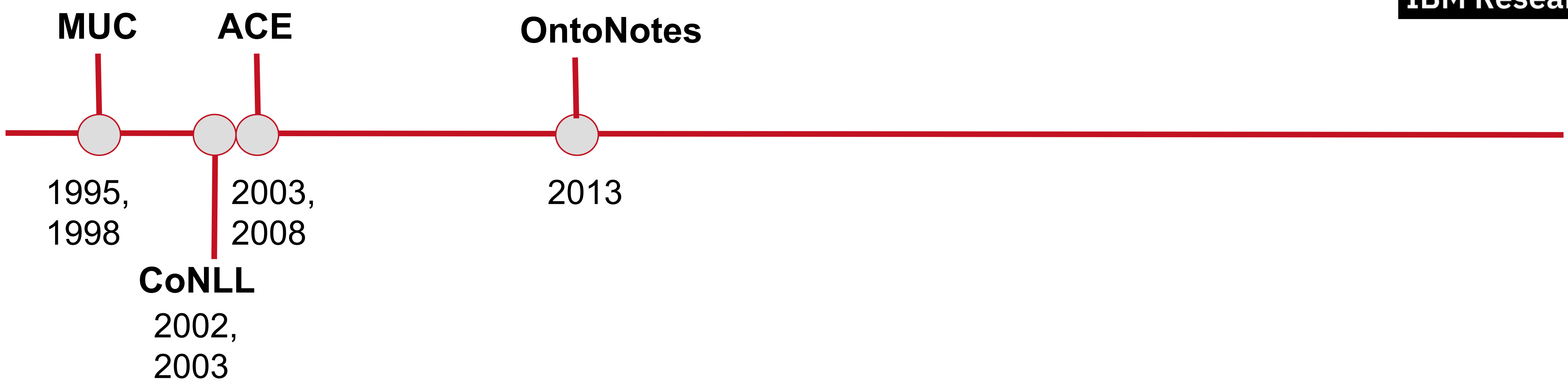


OntoNotes Corpus ver. 5.0 (also leveraged in CoNLL 2013)

- **Entity Types**. 18 types including **Person**, **NORP, Facility, Organization, GPE, Location, Product, Event, Work of art, Law, Language, Date, Time, Percent, Monez, Quantity, Ordinal,, and Cardinal.**

Corpora

IBM Research AI

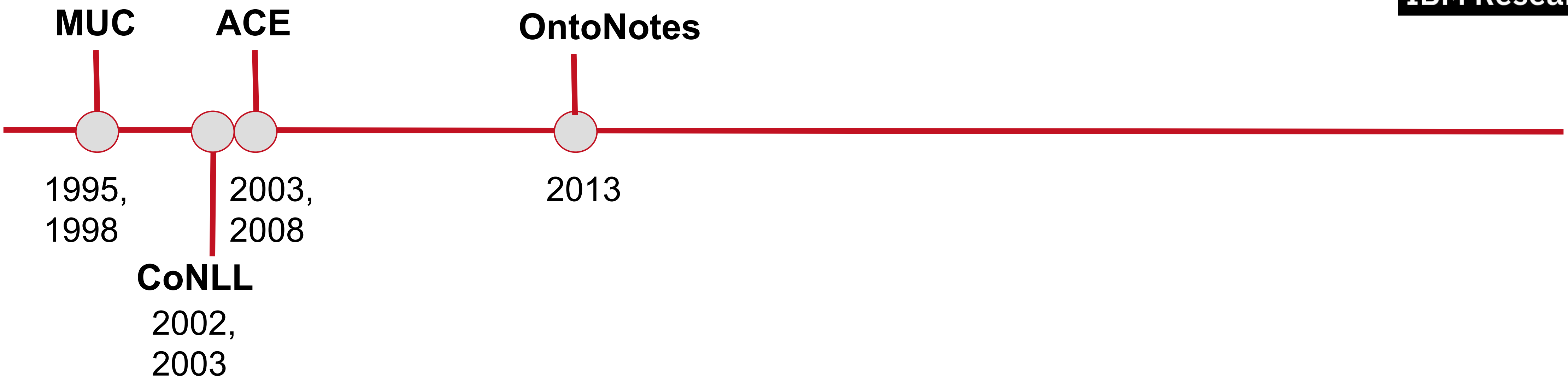


OntoNotes Corpus ver. 5.0 (also leveraged in CoNLL 2013)

- **Entity Types**. 18 types including **Person, NORP, Facility, Organization, GPE, Location, Product, Event, Work of art, Law, Language, Date, Time, Percent, Monez, Quantity, Ordinal,, and Cardinal.**
- **Genre.** telephone conversations, newswire, newsgroups, broadcast news, broadcast conversation, weblogs, religious texts
- **Languages**. English, Arabic, and Chinese

IBM Research AI

Corpora



(Shift in focus) NER based on Encyclopedias

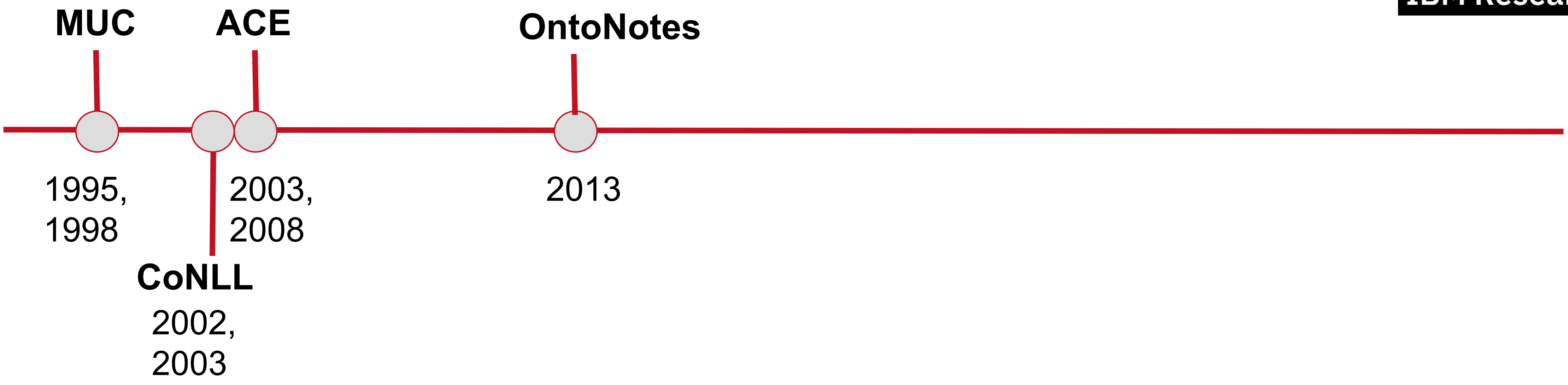
- Encyclopedias (e.g., Wikipedia) are exploited as external knowledge for NER (Kazama and Torisawa, 2007) but corpora for NER are also created based on the large-scale types in Encyclopedias.

Reference

Jun'ichi Kazama and Kentaro Torisawa. "Exploiting Wikipedia as external knowledge for named entity recognition." *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)*. 2007.

IBM Research AI

Corpora

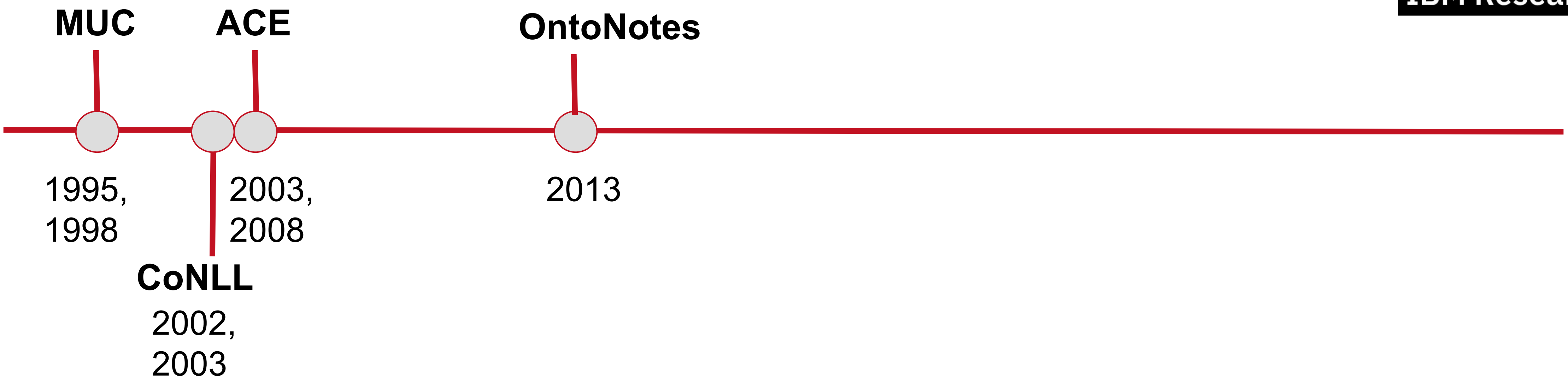


(Shift in focus) NER based on Encyclopedias

- Encyclopedias (e.g., Wikipedia) are exploited as external knowledge for NER but corpora for NER are also created based on the large-scale types in Encyclopedias.

Corpora

IBM Research AI

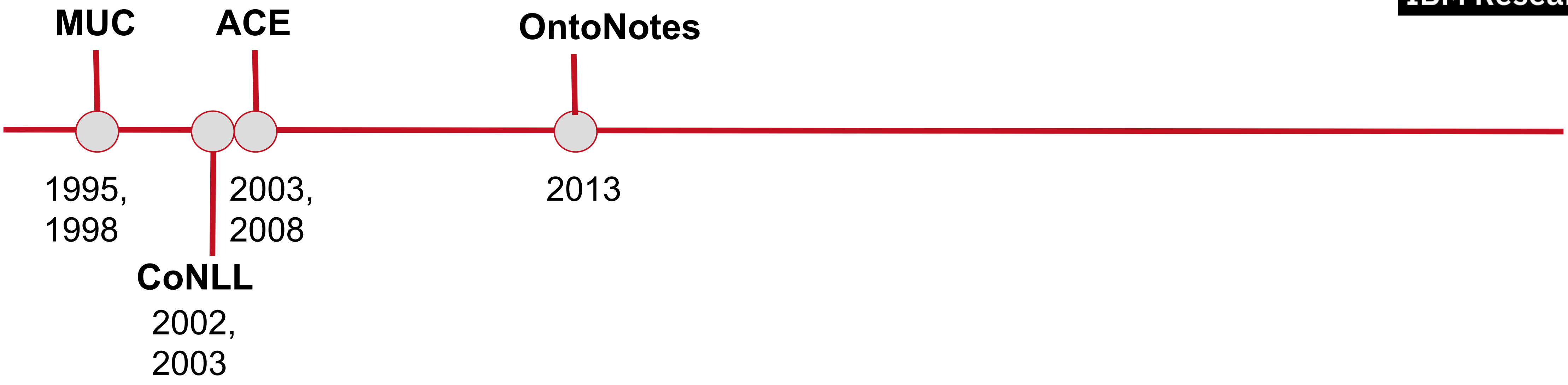


(Shift in focus) NER based on Encyclopedias

- Encyclopedias (e.g., Wikipedia) are exploited as external knowledge for NER but corpora for NER are also created based on the large-scale types in Encyclopedias.
- Facilitates large-scale and automatic type annotation in the range of hundreds of unique types by tapping into hyperlink structures embedded in Encyclopedia web pages.
  - Intuition: elements that are valid candidates for description in an Encyclopedia are described further in a separate webpage and therefore their references in other pages are hyperlinks. E.g., country names, university names, spouse names in pages of public personalities.

Corpora

IBM Research AI

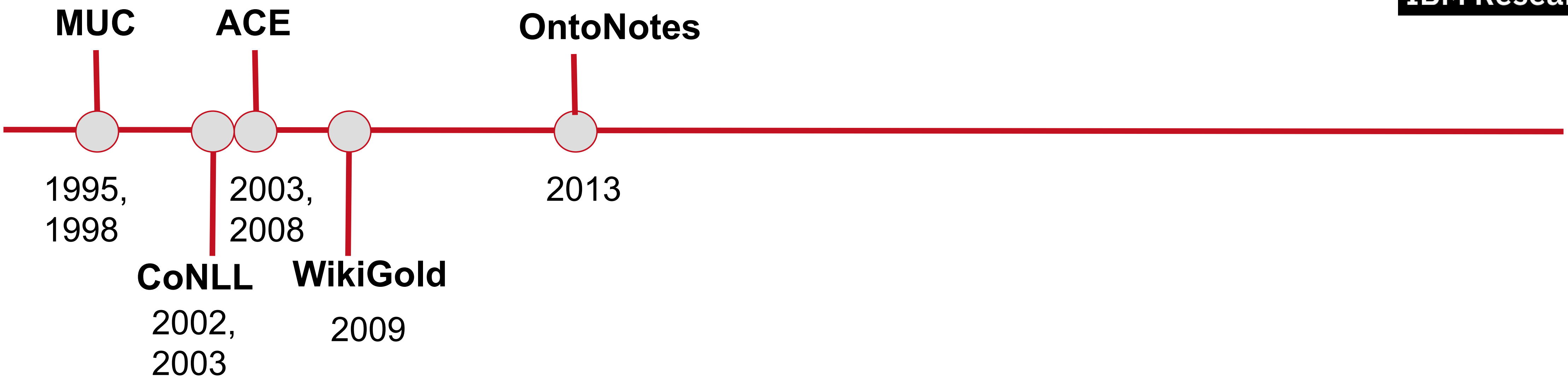


(Shift in focus) NER based on Encyclopedias

- Encyclopedias (e.g., Wikipedia) are exploited as external knowledge for NER but corpora for NER are also created based on the large-scale types in Encyclopedias.
- Facilitates large-scale and automatic type annotation in the range of hundreds of unique types by tapping into hyperlink structures embedded in Encyclopedia web pages.
  - Intuition: elements that are valid candidates for description in an Encyclopedia are described further in a separate webpage and therefore their references in other pages are hyperlinks. E.g., country names, university names, spouse names in pages of public personalities.
- Facilitates relatively easy multilingual corpus generation in addition to obtaining the large-scale named entity types.

Corpora

IBM Research AI



Wikipedia-based NER

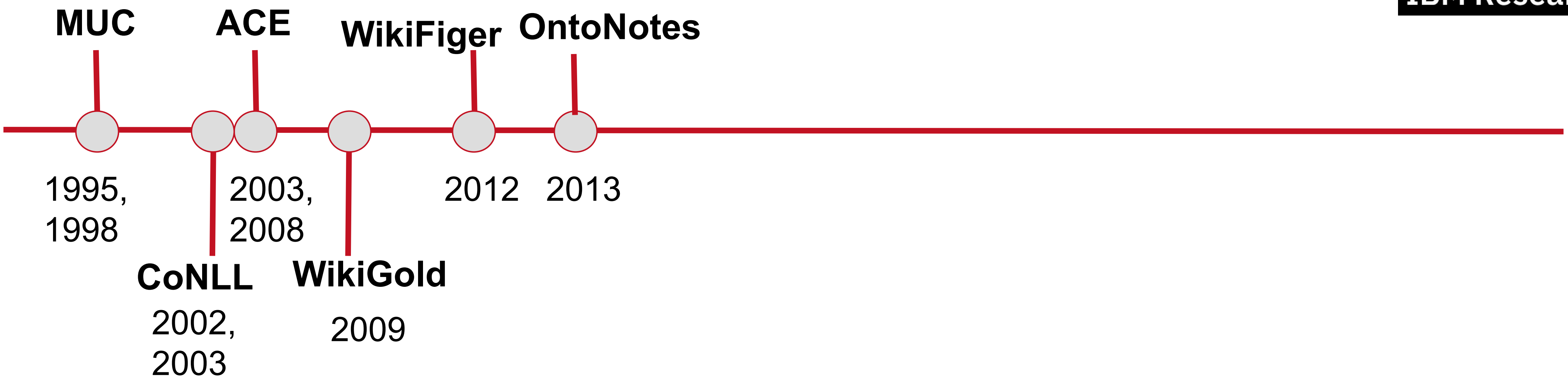
- **WikiGold (WG).** WG is a manually annotated dataset of Wikipedia articles with coarse-grained named entity tags same as the CoNLL types, i.e. Person, Location, Organization, and Miscellaneous. Using Wikipedia's link structure to automatically generate near gold-standard annotations.

Reference

Balasuriya, Dominic, et al. "Named entity recognition in wikipedia." *Proceedings of the 2009 workshop on the people's web meets NLP: Collaboratively constructed semantic resources (People's Web)*. 2009.

Corpora

IBM Research AI



Wikipedia-based NER

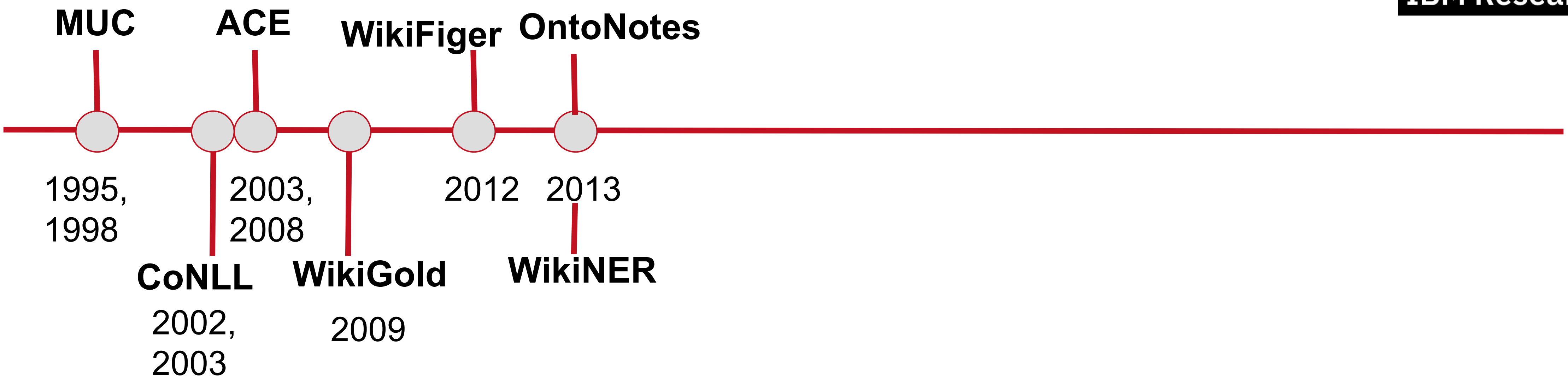
- **WikiFiger.** Curated a set of 112 unique tags based on Freebase types for NER annotations. Some coarse-grained categories are - art, building, event, location, mixed, organization, person, and product.
  - Examples of fine-grained types. Person, for instance, includes Actor, Architect, Artist, Athlete, Author, Coach, Director, Engineer, etc.

Reference

Ling, Xiao, and Daniel S. Weld. "Fine-grained entity recognition." *Twenty-Sixth AAAI Conference on Artificial Intelligence*. 2012.

Corpora

IBM Research AI



Wikipedia-based NER

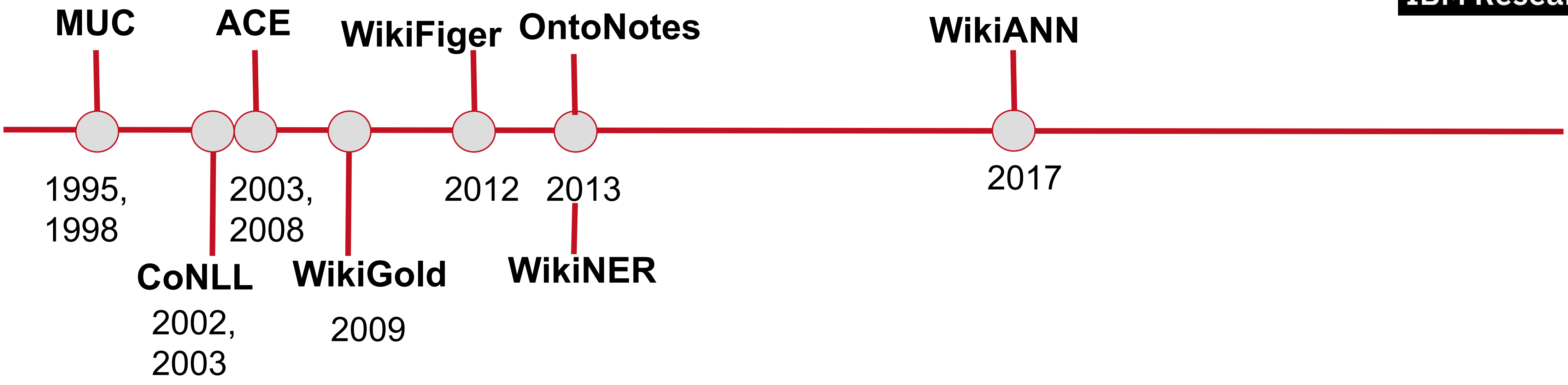
- **WikiNER.** For the first-time, automatically created <u>enormous</u> **multilingual** silver-standard training annotations for named entity recognition (ner) by exploiting the text and structure of Wikipedia across the well-represented languages for the four CoNLL entity types.
  - Languages covered: Dutch, English, French, German, Italian, Polish, Portuguese, Russian, and Spanish

Reference

Nothman, Joel, et al. "Learning multilingual named entity recognition from Wikipedia." *Artificial Intelligence* 194 (2013): 151-175.

Corpora

IBM Research AI



Wikipedia-based NER

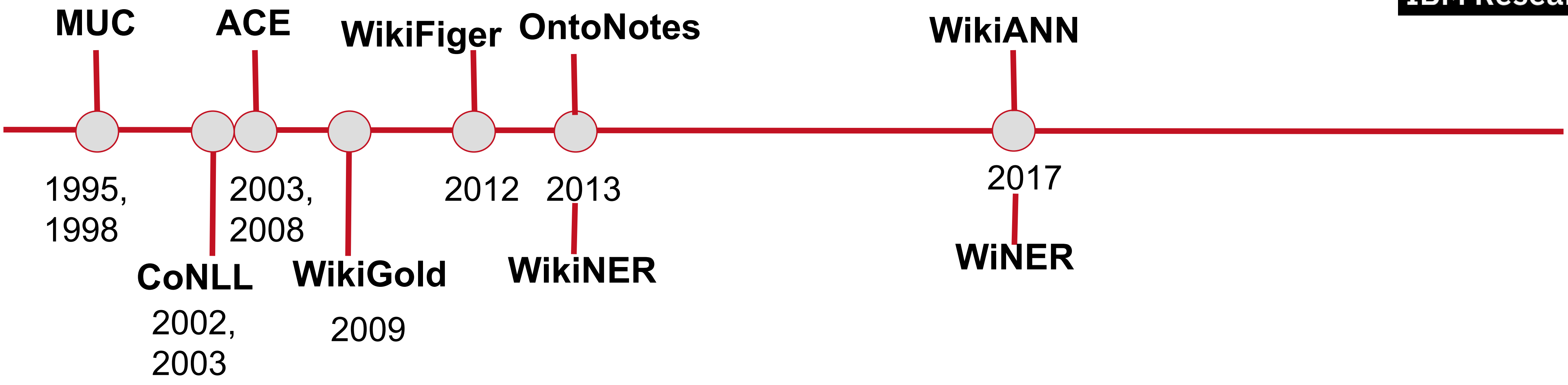
- **WikiANN.** Extended multilinguality from WikiNER's nine languages coverage to 282 (almost all) languages in Wikipedia. Further, incorporated annotations for fine-grained entities including the 139 types in the Abstract Meaning Representation corpus.

Reference

Pan, Xiaoman, et al. "Cross-lingual name tagging and linking for 282 languages." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2017.

Corpora

IBM Research AI



Wikipedia-based NER

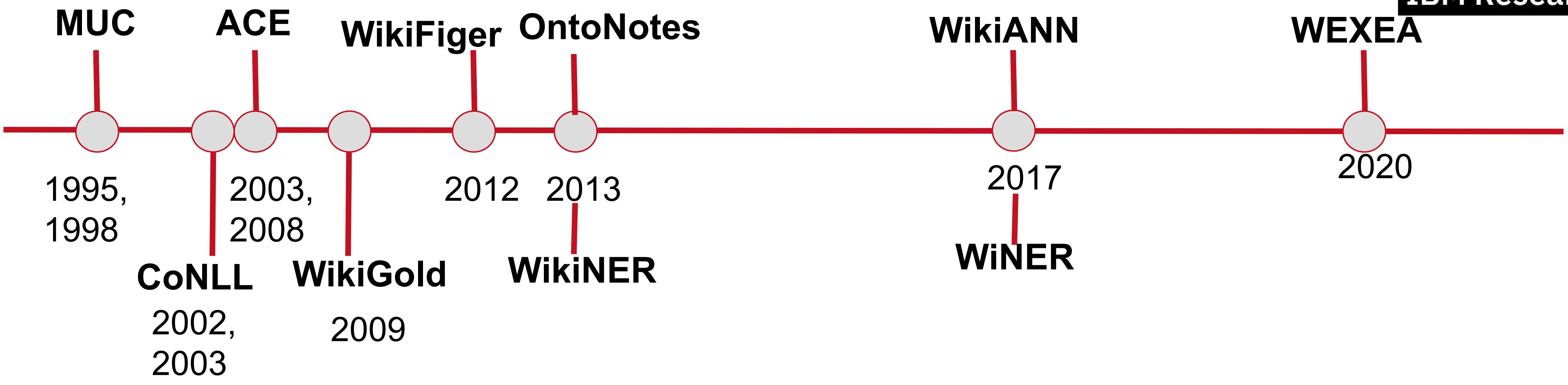
- **WiNER.** Extended the coverage of named entity mentions with the help of coreference resolution.

Reference

Ghaddar, Abbas, and Philippe Langlais. "Winer: A wikipedia annotated corpus for named entity recognition." *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 2017.

Corpora

IBM Research AI



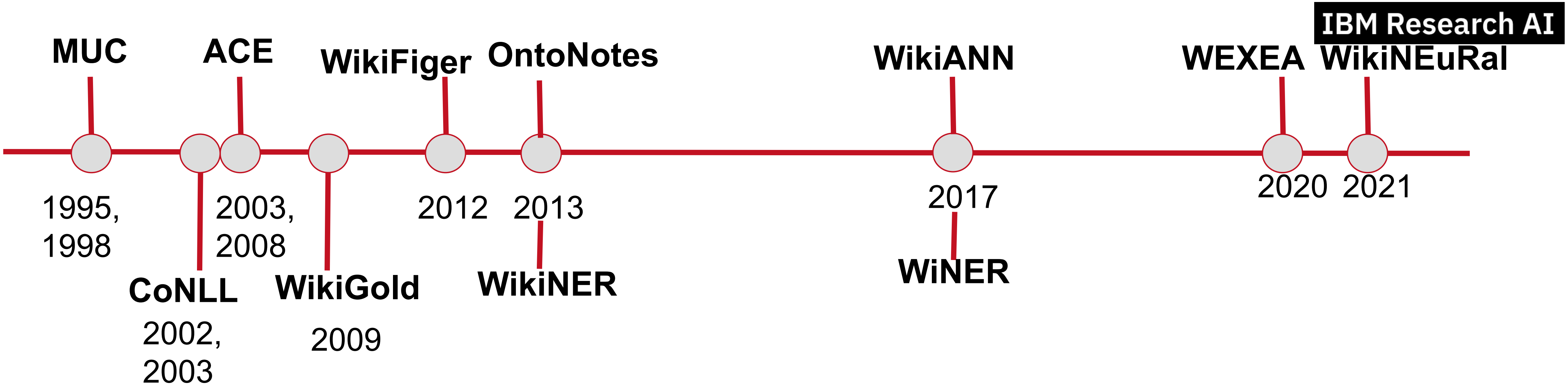
Wikipedia-based NER

- **WEXEA.** Wikipedia EXhaustive Entity Annotation system can create a large annotated corpus based on Wikipedia containing millions of annotations incorporating coreference signals. Further, the system while demonstrated in English Wikipedia can be applied to generate annotations on any language in Wikipedia.

Reference

Strobl, Michael, Amine Trabelsi, and Osmar R. Zaiane. "WEXEA: Wikipedia EXhaustive Entity Annotation." *Proceedings of the 12th Language Resources and Evaluation Conference*. 2020.

Corpora

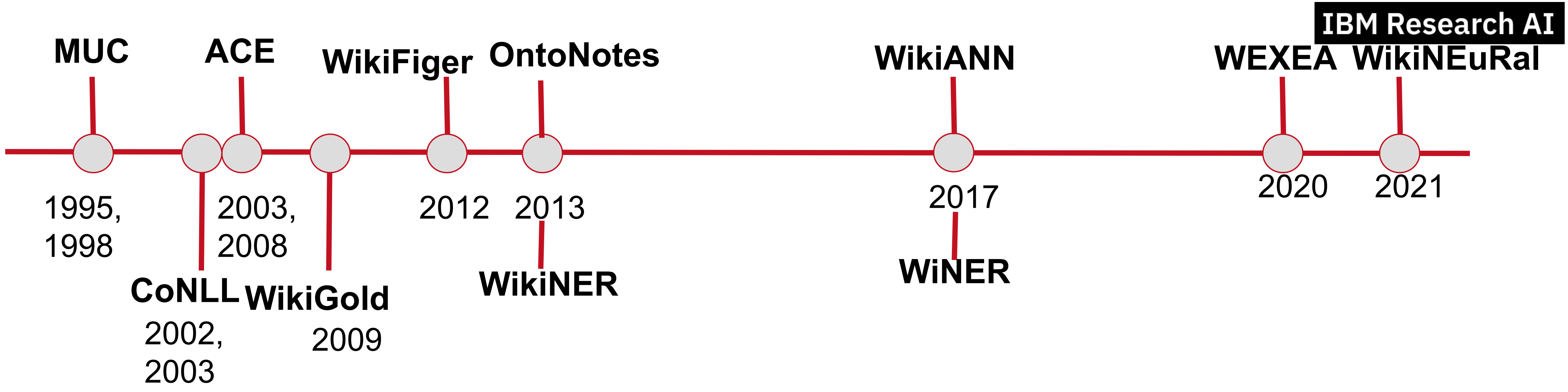


Wikipedia-based NER

- **WikiNEuRal.** Combines neural and knowledge-based method for silver data creation for multilingual NER. Specifically, leverages BabelNet synsets as fine-grained entity types to generate multilingual annotated datasets within a neural annotation framework.

Reference
Tedeschi, Simone, et al. "WikiNEuRal: Combined neural and knowledge-based silver data creation for multilingual NER." *Findings of the Association for Computational Linguistics: EMNLP 2021*. 2021.

Corpora

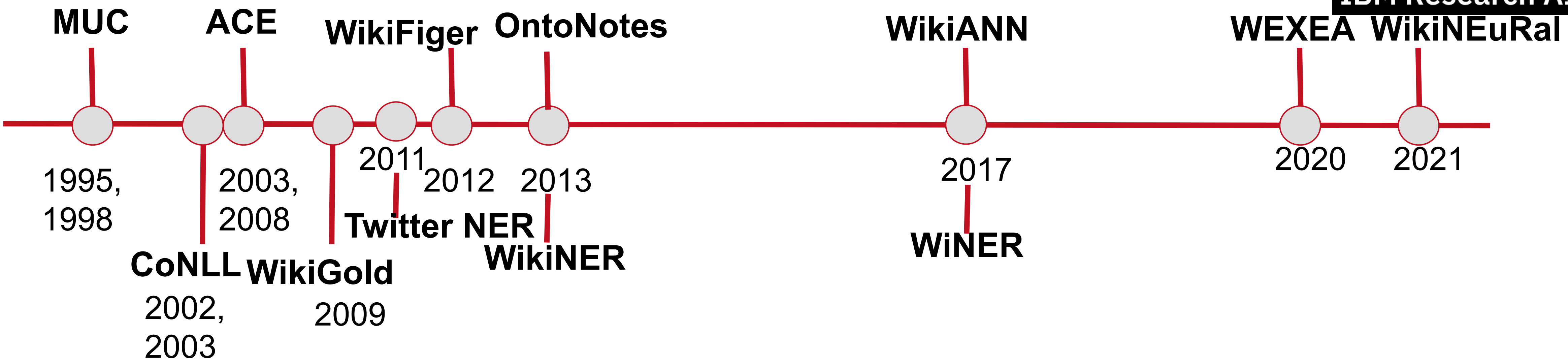


Social media corpus-based NER

- Comprises primarily of tweets
- Twitter is increasingly used in applications to track trending events worldwide. As such information of the entities involved in these events is equally important.
- Communication on social media involve unconventional linguistic particularities in terms of use of emojis and acronyms which is not found in mainstream communication. Consequently, it was noted that the performance of standard NLP tools was severely degraded on tweets. This led to annotation and development of social media or twitter-specific NER corpora and tools.

Corpora

IBM Research AI



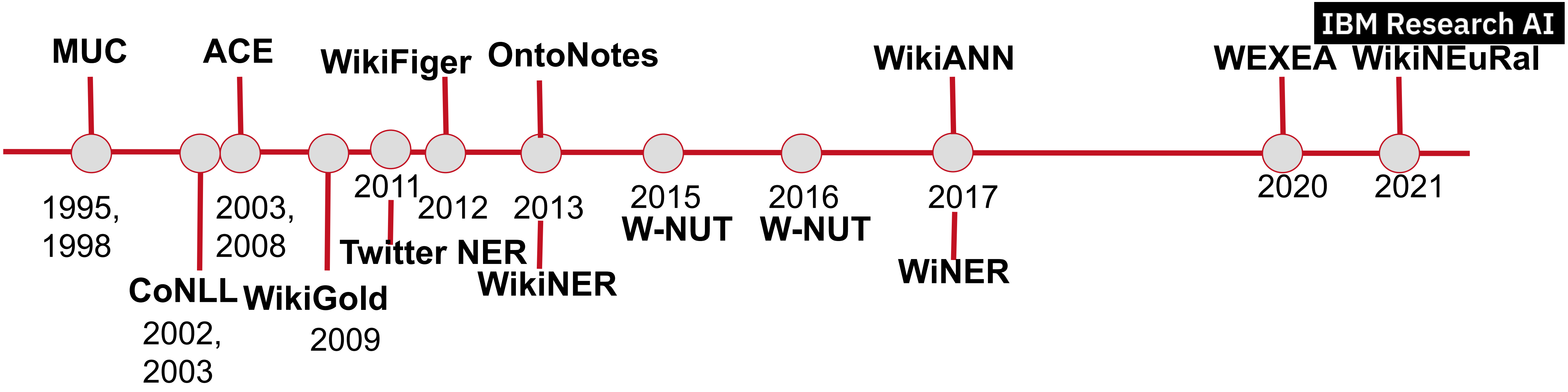
Social media corpus-based NER

- **Twitter NER.** Contains annotations for 2,400 tweets with 10 types which are both popular on Twitter, and have good coverage in Freebase: PERSON, GEO-LOCATION, COMPANY, PRODUCT, FACILITY, TV-SHOW, MOVIE, SPORTSTEAM, BAND, and OTHER. Tool https://github.com/aritter/twitter_nlp

Reference

Ritter, Alan, Sam Clark, and Oren Etzioni. "Named entity recognition in tweets: an experimental study." *Proceedings of the 2011 conference on empirical methods in natural language processing*. 2011.

Corpora

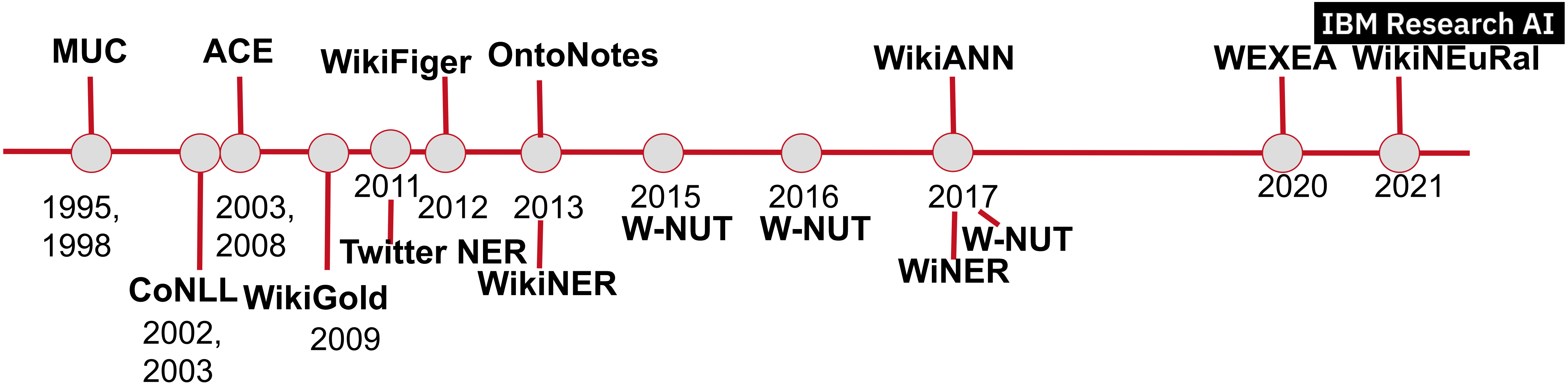


Social media corpus-based NER

- **W-NUT 2015 and 2016.** Extensions of the original Twitter NER corpus with 2400 tweets with 10 types: PERSON, GEO-LOCATION, COMPANY, PRODUCT, FACILITY, TV-SHOW, MOVIE, SPORTSTEAM, MUSIC-ARTIST, and OTHER. Contains additional test dataset annotations.

References
Baldwin, Timothy, et al. "Shared tasks of the 2015 workshop on noisy user-generated text: Twitter lexical normalization and named entity recognition." *Proceedings of the Workshop on Noisy User-generated Text*. 2015.
Strauss, Benjamin, et al. "Results of the wnut16 named entity recognition shared task." *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*. 2016.

Corpora



Social media corpus-based NER

- **W-NUT 2017.** Maintains a specific focus on rare and emerging entities. Annotated six types of entities: corporation, creative-work, group, location, person, and product.

References
Derczynski, Leon, et al. "Results of the WNUT2017 shared task on novel and emerging entity recognition." *Proceedings of the 3rd Workshop on Noisy User-generated Text*. 2017.

IBM Research AI

Other Corpora

- Biomedical NER
  - GENIA corpus (Kim et al., 2003)
    - A semantically annotated corpus for biomedical text mining.
      - Includes annotations for genes, proteins, and other concepts in the Genia ontology.
      - Includes various levels of annotations other than named entities such as POS, syntax, relations, and coreference

References

J.-D. Kim, T. Ohta, Y. Tateisi, J. Tsujii, GENIA corpus—a semantically annotated corpus for bio-textmining, *Bioinformatics*, Volume 19, Issue suppl_1, 3 July 2003, Pages i180–i182, https://doi.org/10.1093/bioinformatics/btg1023

Other Corpora

IBM Research AI

- Biomedical NER
  - GENIA corpus (Kim et al., 2003)
    - A semantically annotated corpus for biomedical text mining.
      - Includes annotations for genes, proteins, and other concepts in the Genia ontology.
      - Includes various levels of annotations other than named entities such as POS, syntax, relations, and coreference

More information: See comprehensive survey on the Open Research Knowledge Graph (ORKG) platform https://www.orkg.org/orkg/comparison/R164231

References

J.-D. Kim, T. Ohta, Y. Tateisi, J. Tsujii, GENIA corpus—a semantically annotated corpus for bio-textmining, *Bioinformatics*, Volume 19, Issue suppl_1, 3 July 2003, Pages i180–i182, https://doi.org/10.1093/bioinformatics/btg1023

Other Corpora

IBM Research AI

- Biomedical and Biochemical NER
  - BioNLP Shared Task Series (2004, 2011, 2013, 2016, 2019)
    - A shared task series organized for biomedical and biochemical text mining over the span of several years including well-known datasets such as GENIA, JNLPBA, Bacteria Biotope, and CRAFT
      - Includes annotations for genes, proteins, bacteria, bacteria locations, drugs, chemical compounds etc.

Other Corpora

IBM Research AI

- Biomedical and Biochemical NER
  - BioNLP Shared Task Series (2004, 2011, 2013, 2016, 2019)
    - A shared task series organized for biomedical and biochemical text mining over the span of several years including well-known datasets such as GENIA, JNLPBA, Bacteria Biotope, and CRAFT
      - Includes annotations for genes, proteins, bacteria, bacteria locations, drugs, chemical compounds etc.

More information: See comprehensive survey article on the Open Research Knowledge Graph (ORKG) platform https://www.orkg.org/orkg/review/R165924

Other Corpora

IBM Research AI

- Biomedical and Biochemical NER
  - BioCreAtIvE Shared Task Series (2005, 2008, 2013, 2014, 2015, 2016, 2021)
    - A shared task series organized for biomedical and biochemical text mining over the span of several years including well-known datasets such as CHEMDNER, Protein Interaction, CEMP, GPRO, DNER etc.
      - Includes annotations for genes, proteins, medications etc.

TIB

IBM Research AI

Other Corpora

- Biomedical and Biochemical NER
  - BioCreAtIvE Shared Task Series (2005, 2008, 2013, 2014, 2015, 2016, 2021)
    - A shared task series organized for biomedical and biochemical text mining over the span of several years including well-known datasets such as CHEMDNER, Protein Interaction, CEMP, GPRO, DNER etc.
      - Includes annotations for genes, proteins, medications etc.

More information: See comprehensive survey article on the Open Research Knowledge Graph (ORKG) platform https://www.orkg.org/orkg/review/R172166

Other Corpora

IBM Research AI

- Scholarly domain-specific NER
  - NER performed on a corpus of scholarly article titles, abstracts or full-text.

Other Corpora

IBM Research AI

- Scholarly domain-specific NER

Corpora	Domain	Coverage	Entity Semantic Types	Size	
				Papers	Entities
FTD	CL	titles, abstracts	focus, domain, technique	426	5,382

Reference
Gupta, Sonal, and Christopher D. Manning. "Analyzing the dynamics of research by extracting key aspects of scientific papers." *Proceedings of 5th international joint conference on natural language processing*. 2011.

Other Corpora

IBM Research AI

- Scholarly domain-specific NER

Corpora	Domain	Coverage	Entity Semantic Types	Size	
				Papers	Entities
FTD	CL	titles, abstracts	focus, domain, technique	426	5,382
ACL-RD-TEC	CL	abstracts	language resource, language resource product, measures and measurements, models, other	300	4,391

Reference
QasemiZadeh, Behrang, and Anne-Kathrin Schumann. "The ACL RD-TEC 2.0: A language resource for evaluating term extraction and entity recognition methods." *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 2016.

Other Corpora

IBM Research AI

- Scholarly domain-specific NER

Corpora	Domain	Coverage	Entity Semantic Types	Size	
				Papers	Entities
FTD	CL	titles, abstracts	focus, domain, technique	426	5,382
ACL-RD-TEC	CL	abstracts	language resource, language resource product, measures and measurements, models, other	300	4,391
ScienceIE	CS, MS, Phy	full text	material, process, task	500	10,994

Reference
Augenstein, Isabelle, et al. "SemEval 2017 Task 10: ScienceIE-Extracting Keyphrases and Relations from Scientific Publications." *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. 2017.

Other Corpora

IBM Research AI

- Scholarly domain-specific NER

Corpora	Domain	Coverage	Entity Semantic Types	Size	
				Papers	Entities
FTD	CL	titles, abstracts	focus, domain, technique	426	5,382
ACL-RD-TEC	CL	abstracts	language resource, language resource product, measures and measurements, models, other	300	4,391
ScienceIE	CS, MS, Phy	full text	material, process, task	500	10,994
SciERC	AI	abstracts	evaluation metric, generic, material, method, task	500	8,089

Reference

Luan, Yi, et al. "Multi-Task Identification of Entities, Relations, and Coreference for Scientific Knowledge Graph Construction." *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018.

Other Corpora

TIB

IBM Research AI

- Scholarly domain-specific NER

Corpora	Domain	Coverage	Entity Semantic Types	Size	
				Papers	Entities
FTD	CL	titles, abstracts	focus, domain, technique	426	5,382
ACL-RD-TEC	CL	abstracts	language resource, language resource product, measures and measurements, models, other	300	4,391
ScienceIE	CS, MS, Phy	full text	material, process, task	500	10,994
SciERC	AI	abstracts	evaluation metric, generic, material, method, task	500	8,089
NLP-TDMS	CL	titles, abstracts, full text	task, dataset, metric, score	332	1,384

Reference

Hou, Yufang, et al. "Identification of Tasks, Datasets, Evaluation Metrics, and Numeric Scores for Scientific Leaderboards Construction." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2019.

Other Corpora

IBM Research AI

- Scholarly domain-specific NER

Corpora	Domain	Coverage	Entity Semantic Types	Size	
				Papers	Entities
FTD	CL	titles, abstracts	focus, domain, technique	426	5,382
ACL-RD-TEC	CL	abstracts	language resource, language resource product, measures and measurements, models, other	300	4,391
ScienceIE	CS, MS, Phy	full text	material, process, task	500	10,994
SciERC	AI	abstracts	evaluation metric, generic, material, method, task	500	8,089
NLP-TDMS	CL	titles, abstracts, full text	task, dataset, metric, score	332	1,384
STEM-ECR	10 STEM disciplines	abstracts	data, material, method, process	110	6,165

Reference

D'Souza, Jennifer, et al. "The STEM-ECR Dataset: Grounding Scientific Entity References in STEM Scholarly Content to Authoritative Encyclopedic and Lexicographic Sources."

Other Corpora

IBM Research AI

- Scholarly domain-specific NER

Corpora	Domain	Coverage	Entity Semantic Types	Size	
				Papers	Entities
FTD	CL	titles, abstracts	focus, domain, technique	426	5,382
ACL-RD-TEC	CL	abstracts	language resource, language resource product, measures and measurements, models, other	300	4,391
ScienceIE	CS, MS, Phy	full text	material, process, task	500	10,994
SciERC	AI	abstracts	evaluation metric, generic, material, method, task	500	8,089
NLP-TDMS	CL	titles, abstracts, full text	task, dataset, metric, score	332	1,384
STEM-ECR	10 STEM disciplines	abstracts	data, material, method, process	110	6,165
...	...	...	...	...	...

For more datasets, see survey in D'Souza, Jennifer, and Sören Auer. "Computer Science Named Entity Recognition in the Open Research Knowledge Graph." arXiv preprint arXiv:2203.14579 (2022).

Conclusions on Named Entities

IBM Research AI

- Despite the various definitions of NE(Named Entity), researchers have reached common consensus on the types of NEs to recognize. We generally divide NEs into two categories:
  - Generic NEs: Person and Location.
  - Domain-specific NEs: proteins, enzymes, genes, methods

Plan for Part I of II of the Talk

- Corpora
- Deep Learning Approaches
- Evaluations and State-of-the-Art

Plan for Part I of II of the Talk

- Corpora
- Deep Learning Approaches
- Evaluations and State-of-the-Art

Approaches to Named Entity Recognition

IBM Research AI

- 4 mainstream approaches used in NER are:
  - **Rule-Based Approaches**: Don’t need annotated data as they rely on hand-crafted rules
  - **Unsupervized Learning Approaches**: Rely on unsupervised algorithms without hand-labelled training examples
  - **Feature-based Supervized Learning**: Rely on supervized algorithms with a lots of feature engineering involved.
  - **Deep Learning Approaches**: Automatically discover representations from raw input

Approaches to Named Entity Recognition

IBM Research AI

- 4 mainstream approaches used in NER are:
    - **Rule-Based Approaches**: Don't need annotated data as they rely on hand-crafted rules
    - **Unsupervized Learning Approaches**: Rely on unsupervised algorithms without hand-labelled training examples
    - **Feature-based Supervized Learning**: Rely on supervized algorithms with a lots of feature engineering involved.
    - **Deep Learning Approaches**: Automatically discover representations from raw input

General Deep Learning Architecture for NER

IBM Research AI

- **Distributed representations for input** consider word- and character-level embeddings as well as the incorporation of additional features.
  - in other words, project words or characters or tokens in a semantic vector space such that they become machine-actionable

General Deep Learning Architecture for NER

IBM Research AI

- **Distributed representations for input** consider word- and character-level embeddings as well as the incorporation of additional features.
    - in other words, project words or characters or tokens in a semantic vector space such that they become machine-actionable
- **Context encoder** is to capture the context dependencies using CNN, RNN, and recently Transformers.

General Deep Learning Architecture for NER

IBM Research AI

- **Distributed representations for input** consider word- and character-level embeddings as well as the incorporation of additional features.
  - in other words, project words or characters or tokens in a semantic vector space such that they become machine-actionable
- **Context encoder** is to capture the context dependencies using CNN, RNN, and recently Transformers.
- **Tag decoder** predicts tags for tokens in the input sentence which can be done with a softmax function or a CRF layer.

General Deep Learning Architecture for NER

IBM Research AI

- **Distributed representations for input** consider word- and character-level embeddings as well as the incorporation of additional features.
  - in other words, project words or characters or tokens in a semantic vector space such that they become machine-actionable
- **Context encoder** is to capture the context dependencies using CNN, RNN, and recently Transformers.
- **Tag decoder** predicts tags for tokens in the input sentence which can be done with a softmax function or a CRF layer.

TIB

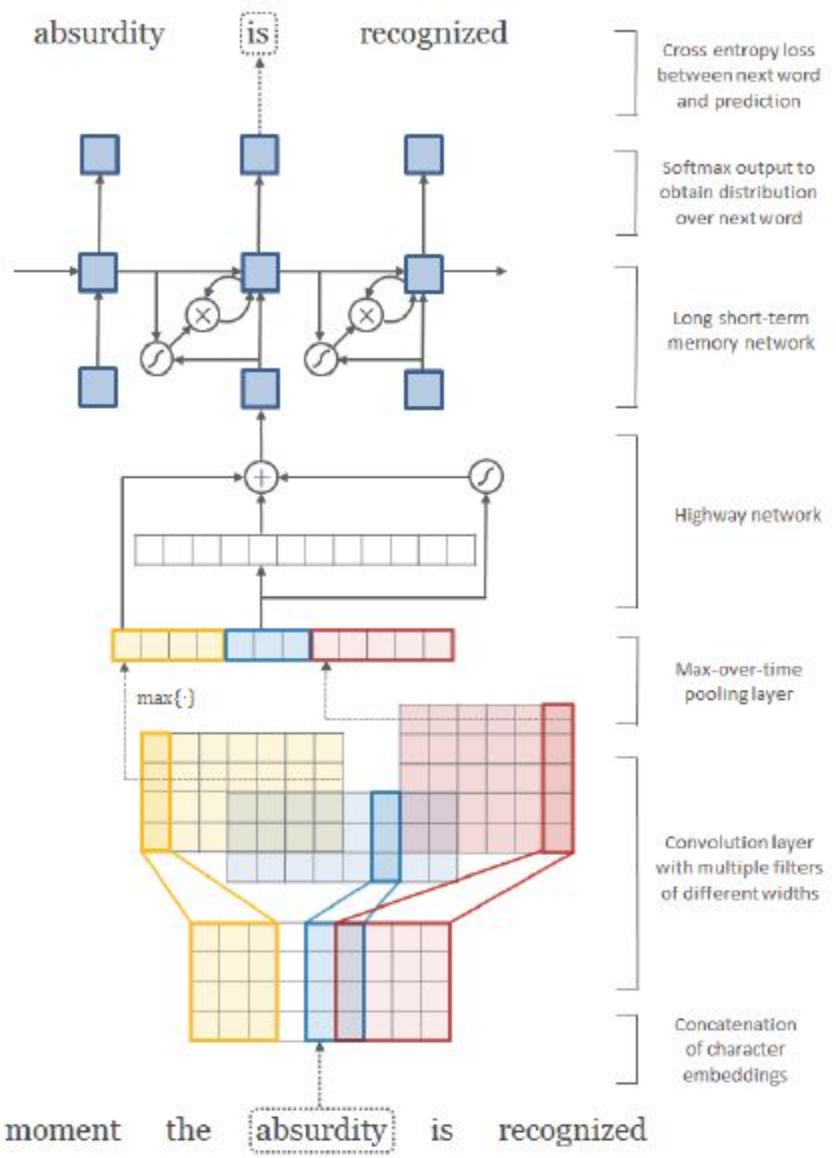
General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Convolutional Neural Networks**
  - Local features around each word are computed with a convolutional layer. After which a global feature vector is constructed by combining the local feature vectors from the convolutional layers. The global feature vectors are then fed to a tag decoder.

General Deep Learning Architecture for NER: Context Encoders

- **Convolutional Neural Networks**
  - Kim et al. (2016) used highway networks over convolution neural networks (CNN) on character sequences of words and then used another layer of LSTM + softmax for the final predictions



Reference

Kim, Yoon, Yacine Jernite, David Sontag, and Alexander M. Rush. "Character-aware neural language models." In *Thirtieth AAAI conference on artificial intelligence*. 2016.

General Deep Learning Architecture for NER: Context Encoders

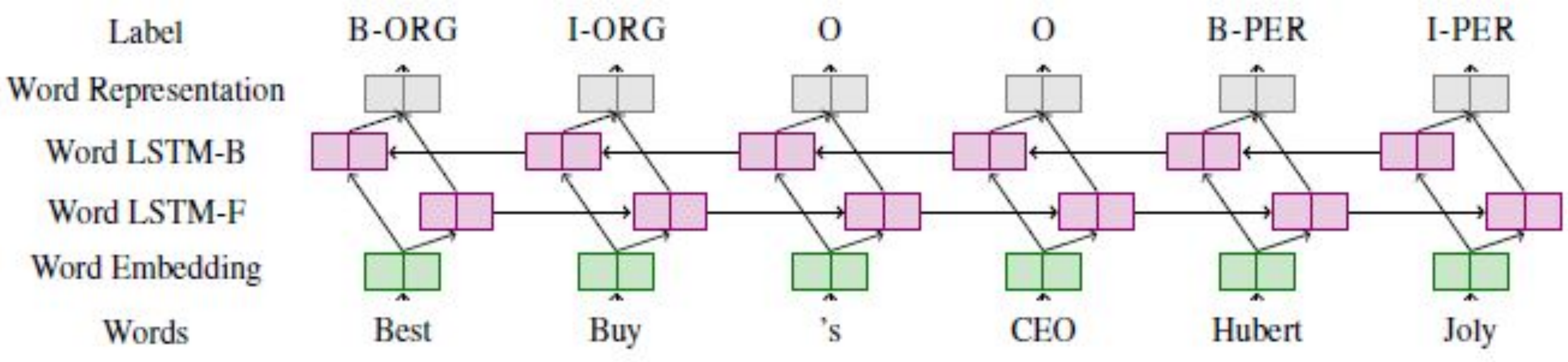
IBM Research AI

- **Recurrent Neural Networks**
  - Bidirectional RNNs such as Bi-LSTMs efficiently make use of past information (via forward states) and future information (via backward states) for a specific time frame. A token encoded by a bidirectional RNN will contain evidence from the whole input sentence.

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Recurrent Neural Networks**
  - Huang et al. (2015) implemented a BiLSTM context encoder that operated over word embeddings and added a CRF layer as a tag decoder



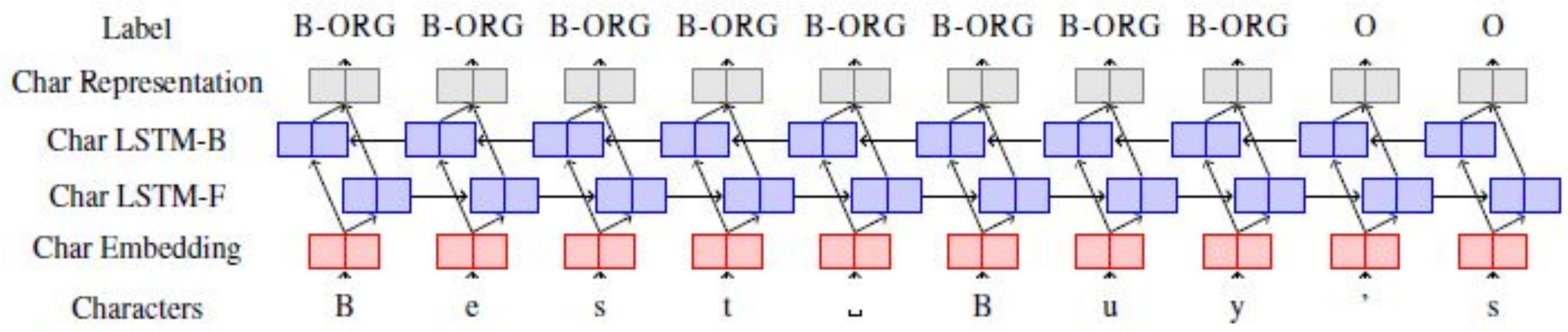
Reference

Huang, Zhiheng, Wei Xu, and Kai Yu. "Bidirectional LSTM-CRF models for sequence tagging." *arXiv preprint arXiv:1508.01991* (2015).

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Recurrent Neural Networks**
  - Kim et al. (2016) implemented a BiLSTM context encoder that operated over character embeddings. In an additional layer before the tag decoder, character labels were transformed into word labels via post processing

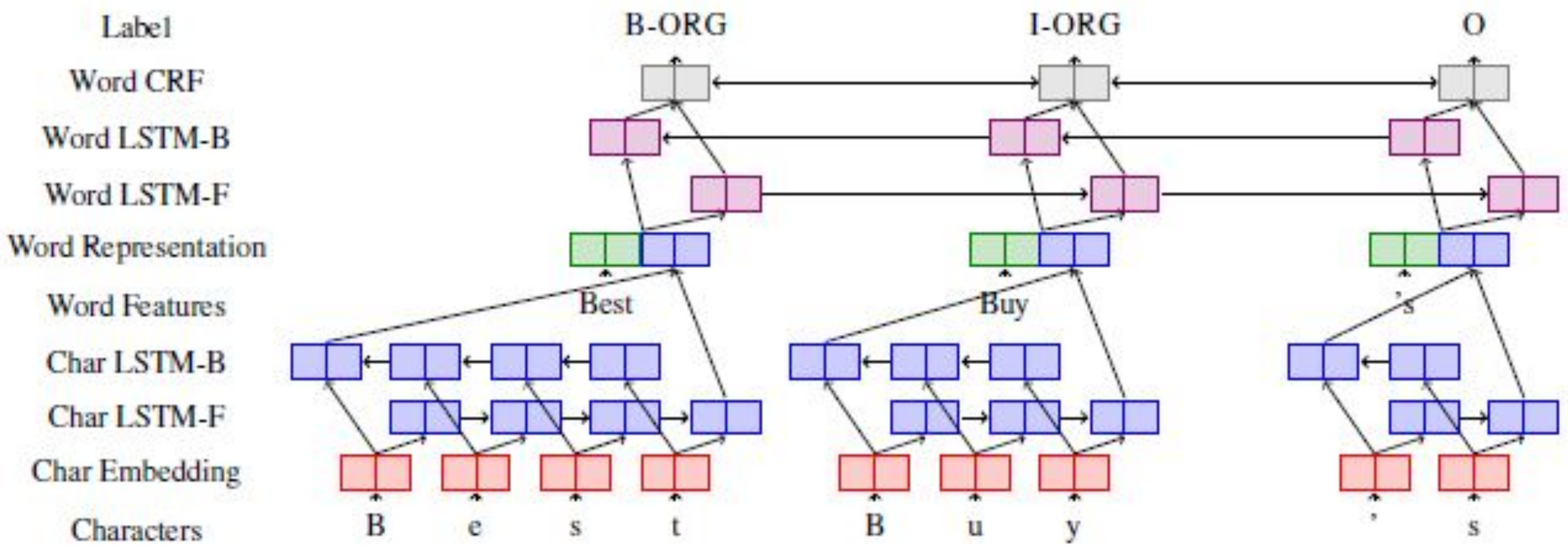


Reference
Kim, Yoon, Yacine Jernite, David Sontag, and Alexander M. Rush. "Character-aware neural language models." In *Thirtieth AAAI conference on artificial intelligence*. 2016.

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Recurrent Neural Networks**
  - Lample et al. (2016) implemented a BiLSTM context encoder that operated over character and word embeddings which finally uses a softmax or CRF layer as the tag decoder.



Reference

Lample, Guillaume, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. "Neural Architectures for Named Entity Recognition." In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 260-270. 2016.

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer**
  - Transformers proposed by Vaswani et al. (2017) dispenses with recurrence and convolutions entirely. A transformer utilizes stacked self-attention and pointwise, fully connected layers to build basic block for encoder and decoder.
  - While RNNs and CNNs have shown promising results, they are affected by long sequence lengths. Instead, transformers relies entirely on attention mechanisms to draw global dependencies between input and output sequences. Thus they can:
    - reduce the computational complexity per layer
    - parallelize more computations
    - capture long-range dependencies effectively

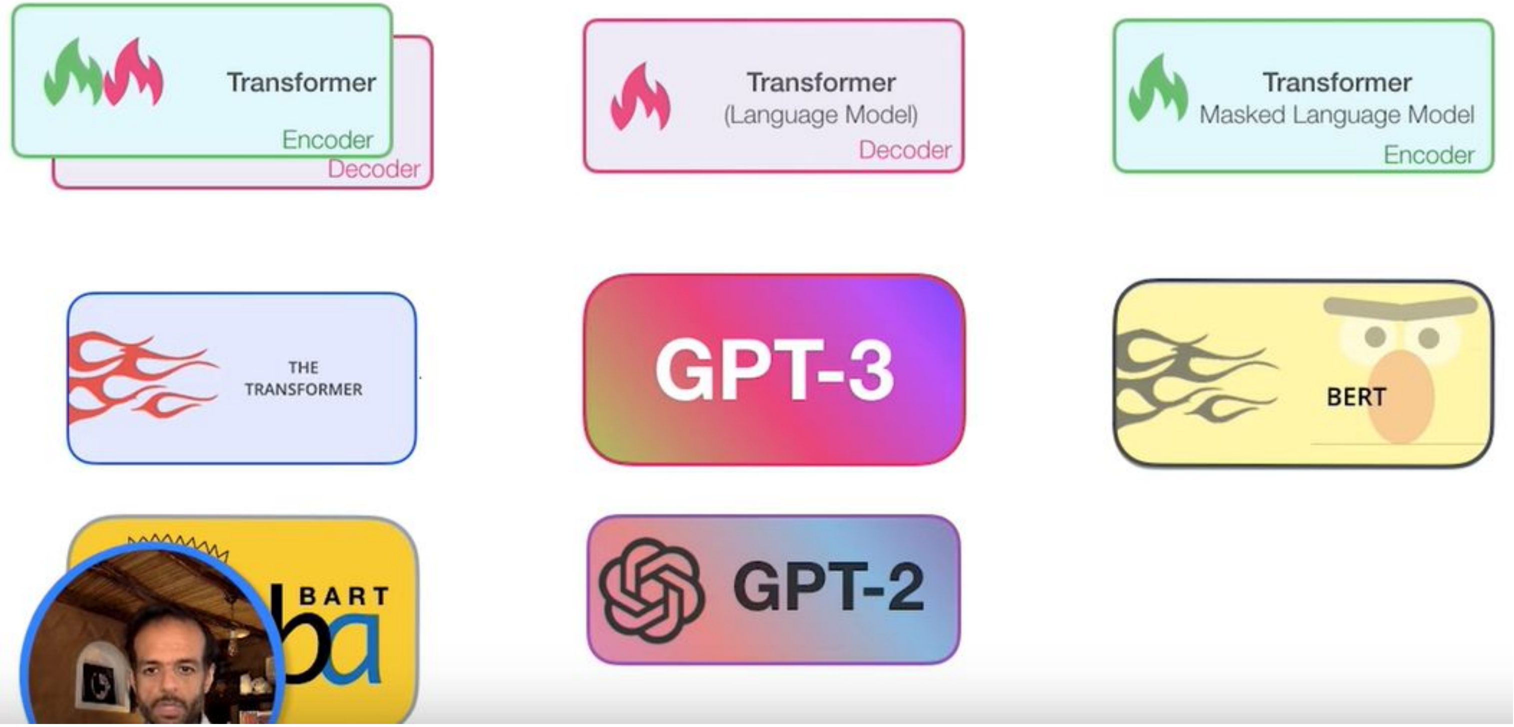
Reference

Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer**
  - Transformers are the basic building for recent Large Language Models (LLMs). There are three main kinds.



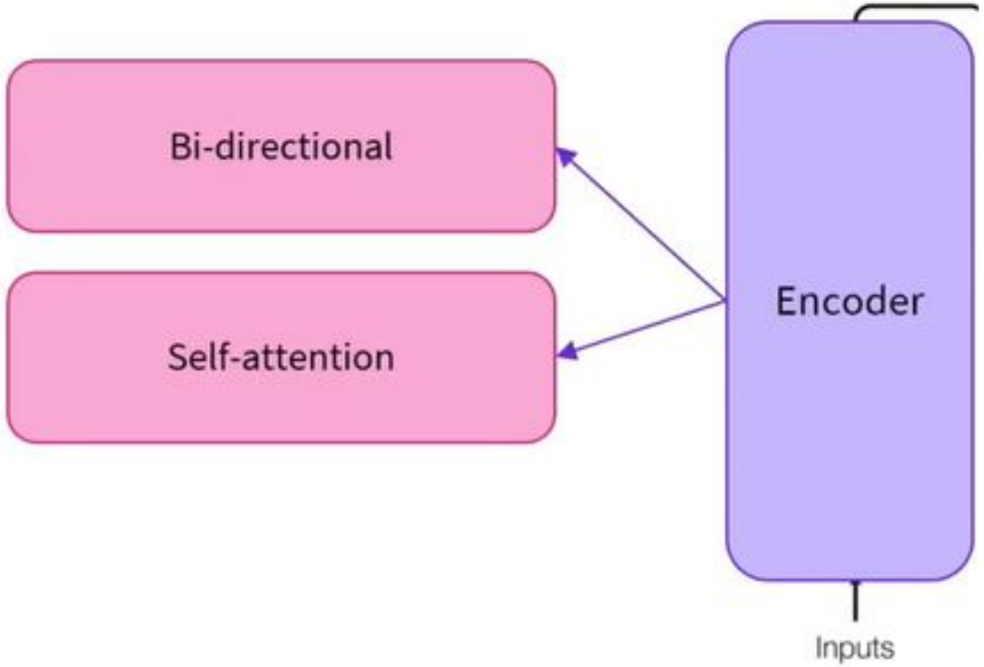
Source

The Narrated Transformer Language Model by Jay Alammar https://www.youtube.com/watch?v=-QH8fRhqFHM&t=1715s

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Encoder only architecture**
    - A single component in an encoder consists of two sub-layers: a multi-head self-attention mechanism and a simple Feed-Forward Network. An encoder can be built from multiple such components.



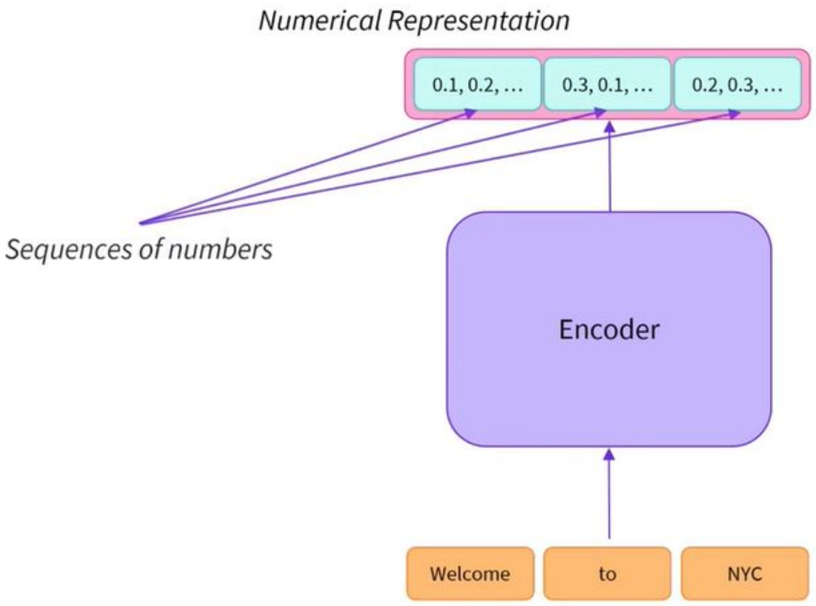
Source

HuggingFace Tutorials https://www.youtube.com/watch?v=MUqNwgPjJvQ

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Encoder only architecture**
  - Outputs a numerical representation as a sequence of numbers in a vector for each word in the input

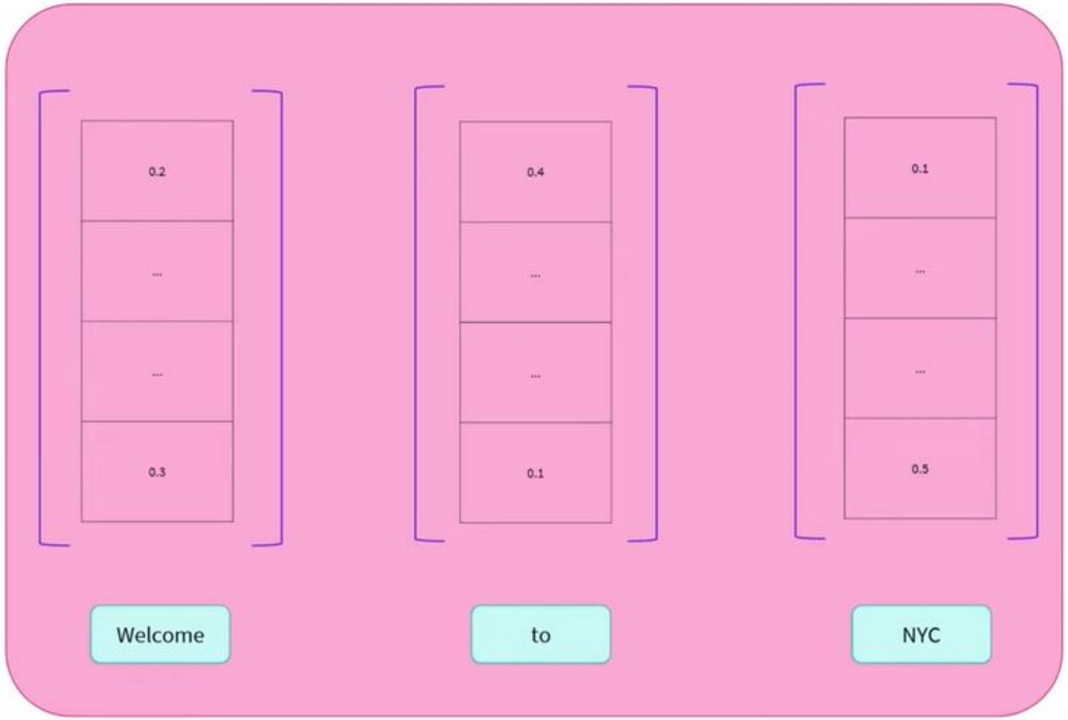


Source
HuggingFace Tutorials https://www.youtube.com/watch?v=MUqNwgPjJvQ

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Encoder only architecture**
  - The dimension of the vector for each word is determined by the architecture of the base model. For BERT-base it is 768.



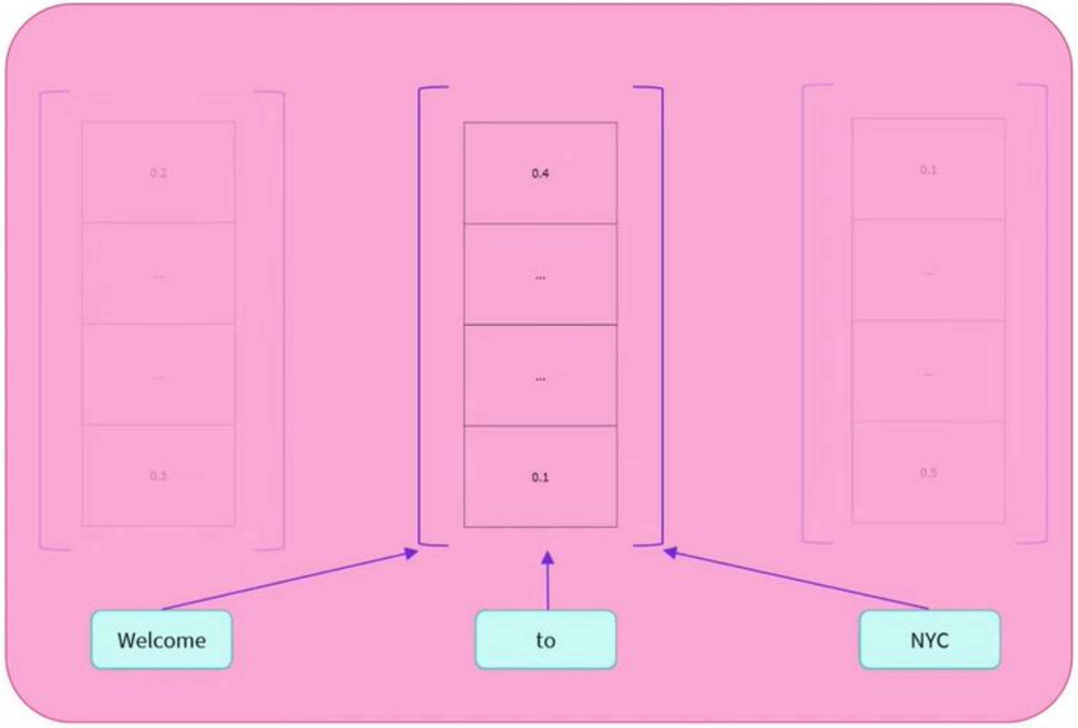
Source

HuggingFace Tutorials https://www.youtube.com/watch?v=MUqNwgPjJvQ

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Encoder only architecture**
  - These representations contain the value of the word but contextualized. Each vector is output based on left and right context of the word. It does this because of self-attention mechanism by which a word sequence is computed weighed on the importance of different parts or words of the input sequence.



Source

HuggingFace Tutorials https://www.youtube.com/watch?v=MUqNwgPjJvQ

General Deep Learning Architecture for NER: Context Encoders

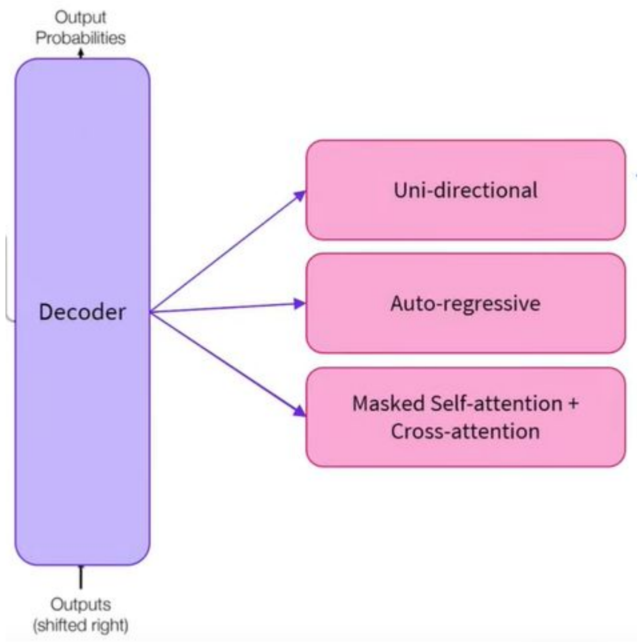
IBM Research AI

- **Transformer - Encoder only architecture**
    - Examples: BERT, RoBERTa, ALBERT

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Decoder only architecture**
    - Like an encoder, a decoder also uses context vectors to compute word sequence representations and attention mechanisms

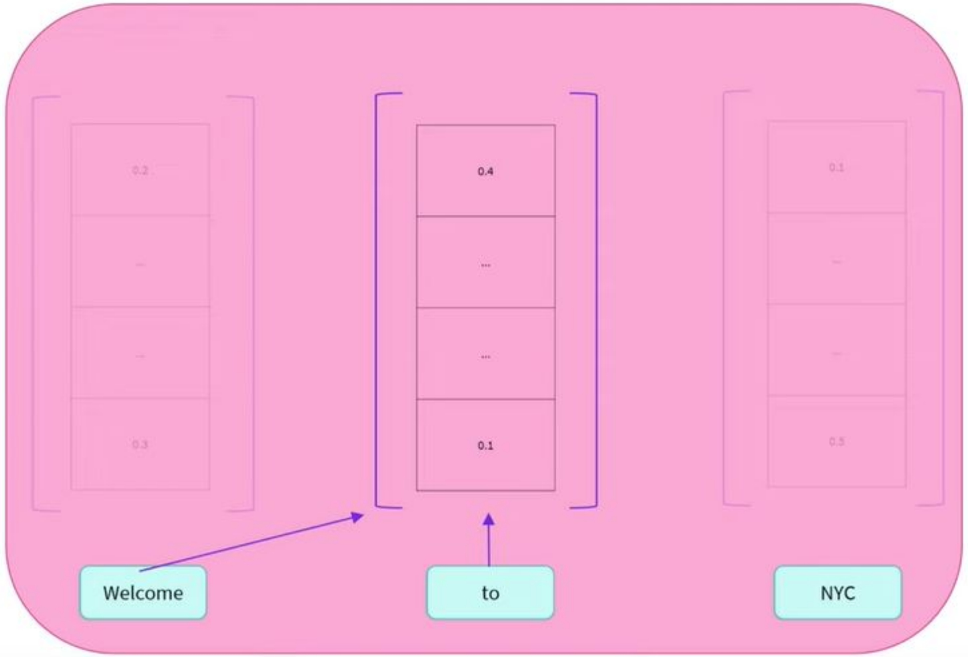


Source
HuggingFace Tutorials https://www.youtube.com/watch?v=d_ixlCubqQw

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Decoder only architecture**
  - Where it differs from an encoder is in its use of the uni-directional masked self-attention mechanism.



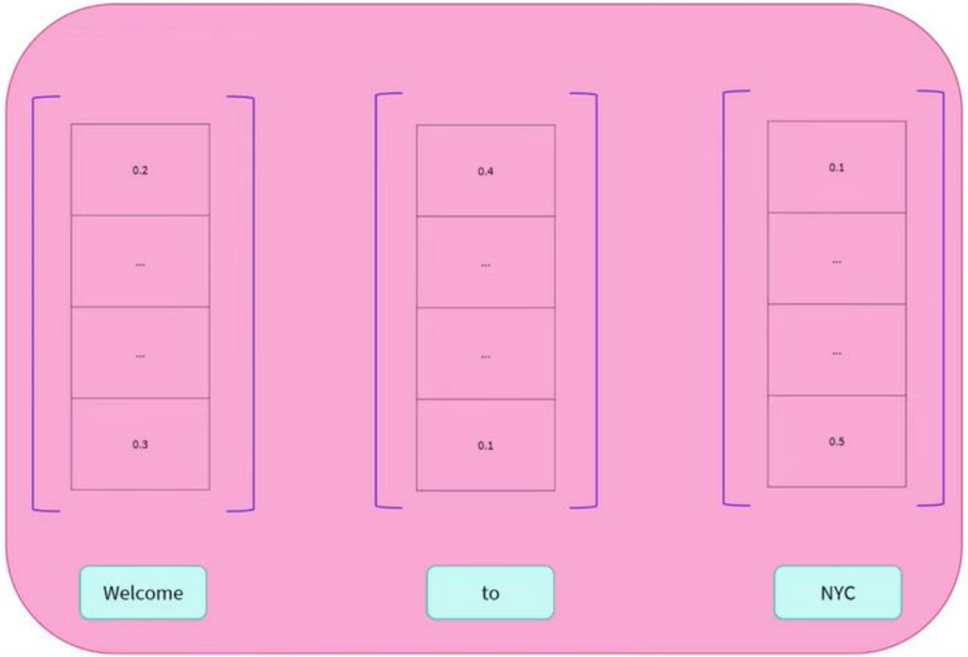
Source

HuggingFace Tutorials https://www.youtube.com/watch?v=d_ixlCubqQw

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Decoder only architecture**
  - The masked self-attention mechanism hides the context vector representation to the left or right of the word.



Source

HuggingFace Tutorials https://www.youtube.com/watch?v=d_ixlCubqQw

General Deep Learning Architecture for NER: Context Encoders

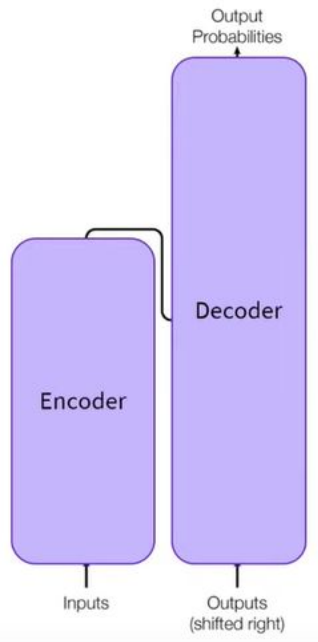
IBM Research AI

- **Transformer - Decoder only architecture**
  - Examples: GPT-3, GPT-Neo

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Encoder-Decoder architecture**
  - Combines the functionality of both the encoder and decoder architectures. Encoders produce models that are very good at natural language understanding. Decoders produce models that are very good at text generation. Encoder-decoders combined produce models that generate text based on functionalities to encode bidirectional contextual sequence representations.



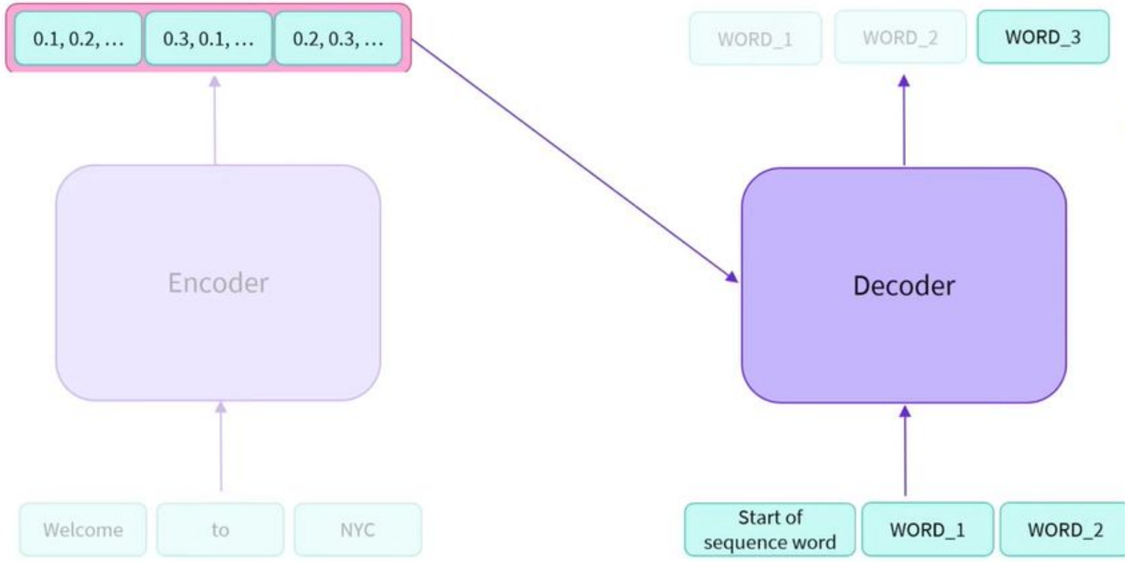
Source

HuggingFace Tutorials https://www.youtube.com/watch?v=0_4KEb08xrE

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Encoder-Decoder architecture**
  - The decoder’s autoregressive behavior allows it to add words that it just generated as output and allows it to include it as part of the generation input sequence.



Source

HuggingFace Tutorials https://www.youtube.com/watch?v=0_4KEb08xrE

General Deep Learning Architecture for NER: Context Encoders

IBM Research AI

- **Transformer - Decoder only architecture**
  - Examples: BART, T5

General Deep Learning Architecture for NER

IBM Research AI

- **Distributed representations for input** consider word- and character-level embeddings as well as the incorporation of additional features.
  - in other words, project words or characters or tokens in a semantic vector space such that they become machine-actionable
- **Context encoder** is to capture the context dependencies using CNN, RNN, or other networks.
- **Tag decoder** predicts tags for tokens in the input sentence which can be done with a softmax function or a CRF layer.

General Deep Learning Architecture for NER

IBM Research AI

Plan for Part I of II of the Talk

- Corpora
- Deep Learning Approaches
- Evaluations and State-of-the-Art

Evaluation on CoNLL 2003 PER, ORG, LOC, and Misc NEs

IBM Research AI

Feature-engineered machine learning systems	Dict	SP	DU	EN	GE
Carreras et al. (2002) binary AdaBoost classifiers	Yes	81.39	77.05	-	-
Malouf (2002) - Maximum Entropy (ME) + features	Yes	73.66	68.08	-	-
Li et al. (2005) SVM with class weights	Yes	-	-	88.3	-
Passos et al. (2014) CRF	Yes	-	-	90.90	-
Ando and Zhang (2005a) Semi-supervised state of the art	No	-	-	89.31	75.27
Agerri and Rigau (2016)	Yes	**84.16**	**85.04**	**91.36**	**76.42**
Feature-inferring neural network word models					
Collobert et al. (2011) Vanilla NN +SLL / Conv-CRF	No	-	-	81.47	-
Huang et al. (2015) Bi-LSTM+CRF	No	-	-	84.26	-
Yan et al. (2016) Win-BiLSTM (English), FF (German) (Many fets)	Yes	-	-	88.91	**76.12**
Collobert et al. (2011) Conv-CRF (SENNA+Gazetteer)	Yes	-	-	89.59	-
Huang et al. (2015) Bi-LSTM+CRF+ (SENNA+Gazetteer)	Yes	-	-	**90.10**	-
Feature-inferring neural network character models					
Gillick et al. (2015) – BTS	No	**82.95**	**82.84**	**86.50**	**76.22**
Kuru et al. (2016) CharNER	No	82.18	79.36	84.52	70.12
Feature-inferring neural network word + character models					
Yang et al. (2017)	Yes	85.77	**85.19**	91.26	-
Luo (2015)	Yes	-	-	91.20	-
Chiu and Nichols (2015)	Yes	-	-	**91.62**	-
Ma and Hovy (2016)	No	-	-	91.21	-
Santos and Guimaraes (2015)	No	82.21	-	-	-
Lample et al. (2016)	No	85.75	81.74	90.94	**78.76**
Bharadwaj et al. (2016)	Yes	**85.81**	-	-	-
Dernoncourt et al. (2017)	No	-	-	90.5	-
Feature-inferring neural network word + character + affix models					
Re-implementation of Lample et al. (2016) (100 Epochs)	No	85.34	85.27	90.24	78.44
Yadav et al. (2018)(100 Epochs)	No	86.92	87.50	90.69	78.56
Yadav et al. (2018) (150 Epochs)	No	**87.26**	**87.54**	90.86	**79.01**

Table from the excellent survey by
Yadav, Vikas, and Steven Bethard. "A Survey on Recent Advances in Named Entity Recognition from Deep Learning models." *Proceedings of the 27th International Conference on Computational Linguistics*. 2018.

Evaluation on CoNLL 2003 PER, ORG, LOC, and Misc NEs

IBM Research AI

Feature-engineered machine learning systems	Dict	SP	DU	EN	GE
Carreras et al. (2002) binary AdaBoost classifiers	Yes	81.39	77.05	-	-
Malouf (2002) - Maximum Entropy (ME) + features	Yes	73.66	68.08	-	-
Li et al. (2005) SVM with class weights	Yes	-	-	88.3	-
Passos et al. (2014) CRF	Yes	-	-	90.90	-
Ando and Zhang (2005a) Semi-supervised state of the art	No	-	-	89.31	75.27
Agerri and Rigau (2016)	Yes	**84.16**	**85.04**	**91.36**	**76.42**
Feature-inferring neural network word models					
Collobert et al. (2011) Vanilla NN +SLL / Conv-CRF	No	-	-	81.47	-
Huang et al. (2015) Bi-LSTM+CRF	No	-	-	84.26	-
Yan et al. (2016) Win-BiLSTM (English), FF (German) (Many fets)	Yes	-	-	88.91	**76.12**
Collobert et al. (2011) Conv-CRF (SENNA+Gazetteer)	Yes	-	-	89.59	-
Huang et al. (2015) Bi-LSTM+CRF+ (SENNA+Gazetteer)	Yes	-	-	**90.10**	-
Feature-inferring neural network character models					
Gillick et al. (2015) – BTS	No	**82.95**	**82.84**	**86.50**	**76.22**
Kuru et al. (2016) CharNER	No	82.18	79.36	84.52	70.12
Feature-inferring neural network word + character models					
Yang et al. (2017)	Yes	85.77	**85.19**	91.26	-
Luo (2015)	Yes	-	-	91.20	-
Chiu and Nichols (2015)	Yes	-	-	**91.62**	-
Ma and Hovy (2016)	No	-	-	91.21	-
Santos and Guimaraes (2015)	No	82.21	-	-	-
Lample et al. (2016)	No	85.75	81.74	90.94	**78.76**
Bharadwaj et al. (2016)	Yes	**85.81**	-	-	-
Dernoncourt et al. (2017)	No	-	-	90.5	-
Feature-inferring neural network word + character + affix models					
Re-implementation of Lample et al. (2016) (100 Epochs)	No	85.34	85.27	90.24	78.44
Yadav et al. (2018)(100 Epochs)	No	86.92	87.50	90.69	78.56
Yadav et al. (2018) (150 Epochs)	No	**87.26**	**87.54**	90.86	**79.01**

Table from the excellent survey by

Yadav, Vikas, and Steven Bethard. "A Survey on Recent Advances in Named Entity Recognition from Deep Learning models." *Proceedings of the 27th International Conference on Computational Linguistics*. 2018.

Evaluation on CoNLL 2003 PER, ORG, LOC, and Misc NEs

Feature-engineered machine learning systems	Dict	SP	DU	EN	GE
Carreras et al. (2002) binary AdaBoost classifiers	Yes	81.39	77.05	-	-
Malouf (2002) - Maximum Entropy (ME) + features	Yes	73.66	68.08	-	-
Li et al. (2005) SVM with class weights	Yes	-	-	88.3	-
Passos et al. (2014) CRF	Yes	-	-	90.90	-
Ando and Zhang (2005a) Semi-supervised state of the art	No	-	-	89.31	75.27
Agerri and Rigau (2016)	Yes	**84.16**	**85.04**	**91.36**	**76.42**
Feature-inferring neural network word models					
Collobert et al. (2011) Vanilla NN +SLL / Conv-CRF	No	-	-	81.47	-
Huang et al. (2015) Bi-LSTM+CRF	No	-	-	84.26	-
Yan et al. (2016) Win-BiLSTM (English), FF (German) (Many fets)	Yes	-	-	88.91	**76.12**
Collobert et al. (2011) Conv-CRF (SENNA+Gazetteer)	Yes	-	-	89.59	-
Huang et al. (2015) Bi-LSTM+CRF+ (SENNA+Gazetteer)	Yes	-	-	**90.10**	-
Feature-inferring neural network character models					
Gillick et al. (2015) – BTS	No	**82.95**	**82.84**	**86.50**	**76.22**
Kuru et al. (2016) CharNER	No	82.18	79.36	84.52	70.12
Feature-inferring neural network word + character models					
Yang et al. (2017)	Yes	85.77	**85.19**	91.26	-
Luo (2015)	Yes	-	-	91.20	-
Chiu and Nichols (2015)	Yes	-	-	**91.62**	-
Ma and Hovy (2016)	No	-	-	91.21	-
Santos and Guimaraes (2015)	No	82.21	-	-	-
Lample et al. (2016)	No	85.75	81.74	90.94	**78.76**
Bharadwaj et al. (2016)	Yes	**85.81**	-	-	-
Dernoncourt et al. (2017)	No	-	-	90.5	-
Feature-inferring neural network word + character + affix models					
Re-implementation of Lample et al. (2016) (100 Epochs)	No	85.34	85.27	90.24	78.44
Yadav et al. (2018)(100 Epochs)	No	86.92	87.50	90.69	78.56
Yadav et al. (2018) (150 Epochs)	No	**87.26**	**87.54**	90.86	**79.01**

Table from the excellent survey by

Yadav, Vikas, and Steven Bethard. "A Survey on Recent Advances in Named Entity Recognition from Deep Learning models." *Proceedings of the 27th International Conference on Computational Linguistics*. 2018.

Evaluation on CoNLL 2003 PER, ORG, LOC, and Misc NEs

IBM Research AI

Feature-engineered machine learning systems	Dict	SP	DU	EN	GE
Carreras et al. (2002) binary AdaBoost classifiers	Yes	81.39	77.05	-	-
Malouf (2002) - Maximum Entropy (ME) + features	Yes	73.66	68.08	-	-
Li et al. (2005) SVM with class weights	Yes	-	-	88.3	-
Passos et al. (2014) CRF	Yes	-	-	90.90	-
Ando and Zhang (2005a) Semi-supervised state of the art	No	-	-	89.31	75.27
Agerri and Rigau (2016)	Yes	**84.16**	**85.04**	**91.36**	**76.42**
Feature-inferring neural network word models					
Collobert et al. (2011) Vanilla NN +SLL / Conv-CRF	No	-	-	81.47	-
Huang et al. (2015) Bi-LSTM+CRF	No	-	-	84.26	-
Yan et al. (2016) Win-BiLSTM (English), FF (German) (Many fets)	Yes	-	-	88.91	**76.12**
Collobert et al. (2011) Conv-CRF (SENNA+Gazetteer)	Yes	-	-	89.59	-
Huang et al. (2015) Bi-LSTM+CRF+ (SENNA+Gazetteer)	Yes	-	-	**90.10**	-
Feature-inferring neural network character models					
Gillick et al. (2015) – BTS	No	**82.95**	**82.84**	**86.50**	**76.22**
Kuru et al. (2016) CharNER	No	82.18	79.36	84.52	70.12
Feature-inferring neural network word + character models					
Yang et al. (2017)	Yes	85.77	**85.19**	91.26	-
Luo (2015)	Yes	-	-	91.20	-
Chiu and Nichols (2015)	Yes	-	-	**91.62**	-
Ma and Hovy (2016)	No	-	-	91.21	-
Santos and Guimaraes (2015)	No	82.21	-	-	-
Lample et al. (2016)	No	85.75	81.74	90.94	**78.76**
Bharadwaj et al. (2016)	Yes	**85.81**	-	-	-
Dernoncourt et al. (2017)	No	-	-	90.5	-
Feature-inferring neural network word + character + affix models					
Re-implementation of Lample et al. (2016) (100 Epochs)	No	85.34	85.27	90.24	78.44
Yadav et al. (2018)(100 Epochs)	No	86.92	87.50	90.69	78.56
Yadav et al. (2018) (150 Epochs)	No	**87.26**	**87.54**	90.86	**79.01**

Table from the excellent survey by
Yadav, Vikas, and Steven Bethard. "A Survey on Recent Advances in Named Entity Recognition from Deep Learning models." *Proceedings of the 27th International Conference on Computational Linguistics*. 2018.

Evaluation on CoNLL 2003 PER, ORG, LOC, and Misc NEs

TIB

IBM Research AI

Feature-engineered machine learning systems	Dict	SP	DU	EN	GE
Carreras et al. (2002) binary AdaBoost classifiers	Yes	81.39	77.05	-	-
Malouf (2002) - Maximum Entropy (ME) + features	Yes	73.66	68.08	-	-
Li et al. (2005) SVM with class weights	Yes	-	-	88.3	-
Passos et al. (2014) CRF	Yes	-	-	90.90	-
Ando and Zhang (2005a) Semi-supervised state of the art	No	-	-	89.31	75.27
Agerri and Rigau (2016)	Yes	**84.16**	**85.04**	**91.36**	**76.42**
Feature-inferring neural network word models					
Collobert et al. (2011) Vanilla NN +SLL / Conv-CRF	No	-	-	81.47	-
Huang et al. (2015) Bi-LSTM+CRF	No	-	-	84.26	-
Yan et al. (2016) Win-BiLSTM (English), FF (German) (Many fets)	Yes	-	-	88.91	**76.12**
Collobert et al. (2011) Conv-CRF (SENNA+Gazetteer)	Yes	-	-	89.59	-
Huang et al. (2015) Bi-LSTM+CRF+ (SENNA+Gazetteer)	Yes	-	-	**90.10**	-
Feature-inferring neural network character models					
Gillick et al. (2015) – BTS	No	**82.95**	**82.84**	**86.50**	**76.22**
Kuru et al. (2016) CharNER	No	82.18	79.36	84.52	70.12
Feature-inferring neural network word + character models					
Yang et al. (2017)	Yes	85.77	**85.19**	91.26	-
Luo (2015)	Yes	-	-	91.20	-
Chiu and Nichols (2015)	Yes	-	-	**91.62**	-
Ma and Hovy (2016)	No	-	-	91.21	-
Santos and Guimaraes (2015)	No	82.21	-	-	-
Lample et al. (2016)	No	85.75	81.74	90.94	**78.76**
Bharadwaj et al. (2016)	Yes	**85.81**	-	-	-
Dernoncourt et al. (2017)	No	-	-	90.5	-
Feature-inferring neural network word + character + affix models					
Re-implementation of Lample et al. (2016) (100 Epochs)	No	85.34	85.27	90.24	78.44
Yadav et al. (2018)(100 Epochs)	No	86.92	87.50	90.69	78.56
Yadav et al. (2018) (150 Epochs)	No	**87.26**	**87.54**	90.86	**79.01**

Table from the excellent survey by

Yadav, Vikas, and Steven Bethard. "A Survey on Recent Advances in Named Entity Recognition from Deep Learning models." *Proceedings of the 27th International Conference on Computational Linguistics*. 2018.

Evaluation on CoNLL 2003 PER, ORG, LOC, and Misc NEs

IBM Research AI

Feature-engineered machine learning systems	Dict	SP	DU	EN	GE
Carreras et al. (2002) binary AdaBoost classifiers	Yes	81.39	77.05	-	-
Malouf (2002) - Maximum Entropy (ME) + features	Yes	73.66	68.08	-	-
Li et al. (2005) SVM with class weights	Yes	-	-	88.3	-
Passos et al. (2014) CRF	Yes	-	-	90.90	-
Ando and Zhang (2005a) Semi-supervised state of the art	No	-	-	89.31	75.27
Agerri and Rigau (2016)	Yes	**84.16**	**85.04**	**91.36**	**76.42**
Feature-inferring neural network word models					
Collobert et al. (2011) Vanilla NN +SLL / Conv-CRF	No	-	-	81.47	-
Huang et al. (2015) Bi-LSTM+CRF	No	-	-	84.26	-
Yan et al. (2016) Win-BiLSTM (English), FF (German) (Many fets)	Yes	-	-	88.91	**76.12**
Collobert et al. (2011) Conv-CRF (SENNA+Gazetteer)	Yes	-	-	89.59	-
Huang et al. (2015) Bi-LSTM+CRF+ (SENNA+Gazetteer)	Yes	-	-	**90.10**	-
Feature-inferring neural network character models					
Gillick et al. (2015) – BTS	No	**82.95**	**82.84**	**86.50**	**76.22**
Kuru et al. (2016) CharNER	No	82.18	79.36	84.52	70.12
Feature-inferring neural network word + character models					
Yang et al. (2017)	Yes	85.77	**85.19**	91.26	-
Luo (2015)	Yes	-	-	91.20	-
Chiu and Nichols (2015)	Yes	-	-	**91.62**	-
Ma and Hovy (2016)	No	-	-	91.21	-
Santos and Guimaraes (2015)	No	82.21	-	-	-
Lample et al. (2016)	No	85.75	81.74	90.94	**78.76**
Bharadwaj et al. (2016)	Yes	**85.81**	-	-	-
Dernoncourt et al. (2017)	No	-	-	90.5	-
Feature-inferring neural network word + character + affix models					
Re-implementation of Lample et al. (2016) (100 Epochs)	No	85.34	85.27	90.24	78.44
Yadav et al. (2018)(100 Epochs)	No	86.92	87.50	90.69	78.56
Yadav et al. (2018) (150 Epochs)	No	**87.26**	**87.54**	90.86	**79.01**

Table from the excellent survey by

Yadav, Vikas, and Steven Bethard. "A Survey on Recent Advances in Named Entity Recognition from Deep Learning models." *Proceedings of the 27th International Conference on Computational Linguistics*. 2018.

Performance evaluations on other datasets…

IBM Research AI

- Recommended resource: Leaderboards on PapersWithCode
  https://paperswithcode.com/

LEIBNIZ-INFORMATIONSZENTRUM
TECHNIK UND NATURWISSENSCHAFTEN
UNIVERSITÄTSBIBLIOTHEK

TIB

IBM Research AI

The State of the Art Transformer Language Models on Knowledge Graph Construction from Text

Part 2: Relation Extraction (RE) Perspective

Presented by: Nandana Mihindukulasooriya,
Senior Research Scientist, IBM Research
https://research.ibm.com/ | @nandanamihindu
IBM Research, Ireland

Leibniz
Gemeinschaft

Relation Extraction

TIB

IBM Research AI

Irene Morgan, who was born and raised in Baltimore, lived on Long Island.

Named Entity Recognition

Irene Morgan [PERSON], who was born and raised in **Baltimore** [PLACE], lived on **Long Island** [PLACE].

Relation Extraction

Irene Morgan, who was born and raised in **Baltimore**, lived on Long Island.

Irene Morgan $\xrightarrow{\text{per:city\_of\_birth}}$ Baltimore

A predefined set of relations

Closely Related Tasks

Open Information Extraction (OpenIE)

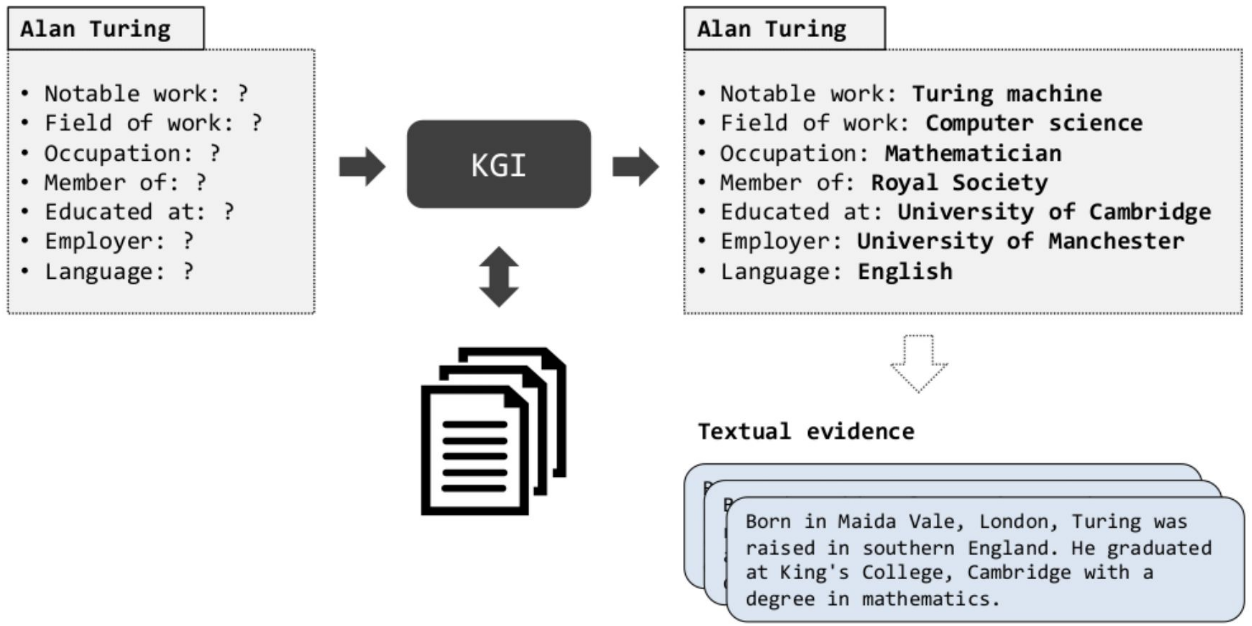
KGC 2020 took place in New York.

(KGC 2020, took place in, New York)

- Relations are not predefined, automatically discovered in text.
- A large number of sparse and diverse relations
- Need to further steps of clustering, canonicalization, alignment to map to a set of KG relations.

Open Information Extraction from the Web. Banko et al. IJACAI 2007.

Slot Filling / Knowledge Base Population



Robust Retrieval Augmented Generation for Zero-shot Slot Filling. Glass et al. EMNLP 2021.

Relation Extraction – Binary Classification

IBM Research AI

[TACRED] Position-aware Attention and Supervised Data Improve Slot Filling. Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, Christopher D. Manning. EMNLP 2017.

Pandit worked at the brokerage Morgan Stanley for about 11 years until 2005, when he and some Morgan Stanley colleagues quit and later founded the hedge fund **Old Lane Partners**.	**Types:** ORGANIZATION/PERSON **Relation:** *org:founded_by*
Baldwin declined further comment, and said JetBlue chief **executive** Dave Barger was unavailable.	**Types:** PERSON/TITLE **Relation:** *no_relation*

Relation Extraction – Multilabel Classification

IBM Research AI

Example	Entity Types & Label
Carey will succeed **Cathleen P. Black**, who held the position for 15 years and will take on a new role as **chairwoman** of Hearst Magazines, the company said.	**Types:** PERSON/TITLE **Relation:** *per:title*
Irene Morgan Kirkaldy, who was born and reared in **Baltimore**, lived on Long Island and ran a child-care center in Queens with her second husband, Stanley Kirkaldy.	**Types:** PERSON/CITY **Relation:** *per:city_of_birth*
Pandit worked at the brokerage Morgan Stanley for about 11 years until 2005, when he and some Morgan Stanley colleagues quit and later founded the hedge fund **Old Lane Partners**.	**Types:** ORGANIZATION/PERSON **Relation:** *org:founded_by*
Baldwin declined further comment, and said JetBlue chief **executive** Dave Barger was unavailable.	**Types:** PERSON/TITLE **Relation:** *no_relation*

[TACRED] Position-aware Attention and Supervised Data Improve Slot Filling. Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, Christopher D. Manning. EMNLP 2017.

Multilabel Classification – different types of relations

IBM Research AI

Cause-Effect	*Smoking* causes *cancer*.
Instrument-Agency	The *murderer* used an *axe*.
Product-Producer	*Bees* make *honey*.
Content-Container	The *cat* is in the *hat*.
Entity-Origin	*Vinegar* is made from *wine*.
Entity-Destination	The *car* arrived at the *station*.
Component-Whole	The *laptop* has a *fast processor*.
Member-Collection	There are ten *cows* in the *herd*.
Communication-Topic	You interrupted a *lecture* on *maths*.

People in Hawaii might be feeling <e1>aftershocks</e1> from that powerful <e2>earthquake</e2> for weeks.

Relation: Cause-Effect(e1, e2)

[SemEval-2010 Task 8] SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals. Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, Stan Szpakowicz. SemEval - ACL 2010.

Document-level RE (DocRED)

IBM Research AI

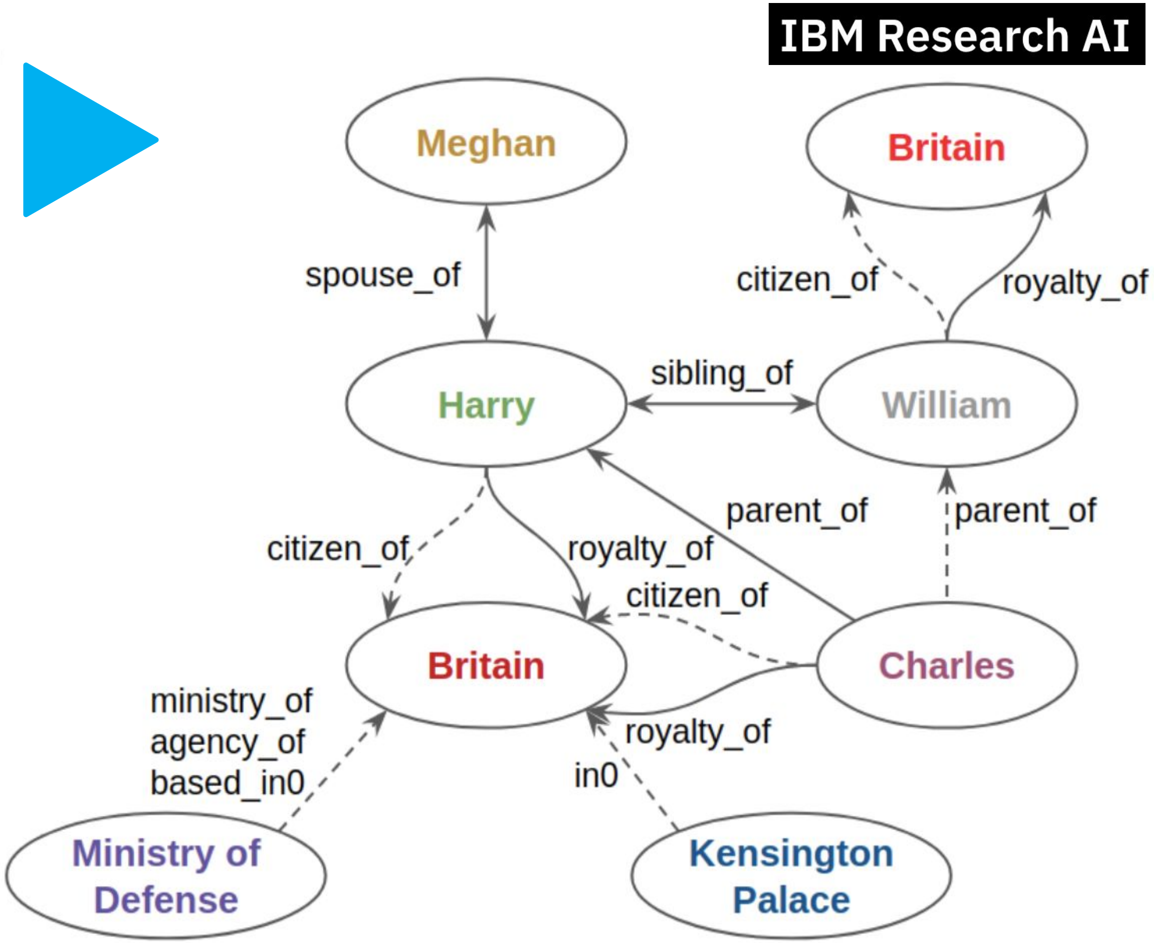
Reasoning Types	%	Examples
Pattern recognition	38.9	[1] *Me Musical Nephews* is a 1942 one-reel animated cartoon directed by Seymour Kneitel and animated by Tom Johnson and George Germanetti. [2] Jack Mercer and Jack Ward wrote the script. ... **Relation:** publication_date **Supporting Evidence:** 1
Logical reasoning	26.6	[1] "Nisei" is the ninth episode of the third season of the American science fiction television series The X-Files. ... [3] It was directed by David Nutter, and written by Chris Carter, Frank Spotnitz and Howard Gordon. ... [8] The show centers on FBI special agents *Fox Mulder* (David Duchovny) and Dana Scully (Gillian Anderson) who work on cases linked to the paranormal, called X-Files. ... **Relation:** creator **Supporting Evidence:** 1, 3, 8
Coreference reasoning	17.6	[1] *Dwight Tillery* is an American politician of the Democratic Party who is active in local politics of Cincinnati, Ohio. ... [3] He also holds a law degree from the University of Michigan Law School. [4] *Tillery* served as mayor of Cincinnati from 1991 to 1993. **Relation:** educated_at **Supporting Evidence:** 1, 3
Common-sense reasoning	16.6	[1] *William Busac* (1020-1076), son of William I, Count of Eu, and his wife Lesceline. ... [4] *William* appealed to King Henry I of France, who gave him in marriage Adelaide, the heiress of the county of Soissons. [5] Adelaide was daughter of Renaud I, Count of Soissons, and Grand Master of the Hotel de France. ... [7] *William* and Adelaide had four children: ... **Relation:** spouse **Supporting Evidence:** 4, 7

[DocRED] DocRED: A Large-Scale Document-Level Relation Extraction Dataset. Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zheng-Hao Liu, Zhiyuan Liu, Lixin Huang, Jie zhou, Maosong Sun. ACL 2019.

From Relation Extraction to Knowledge Graphs

Some additional steps are required

- Entity clustering / canonicalization
- Entity resolution
- Entity linking
  - (adding new entities if necessary)
- Schema matching
- Relation linking

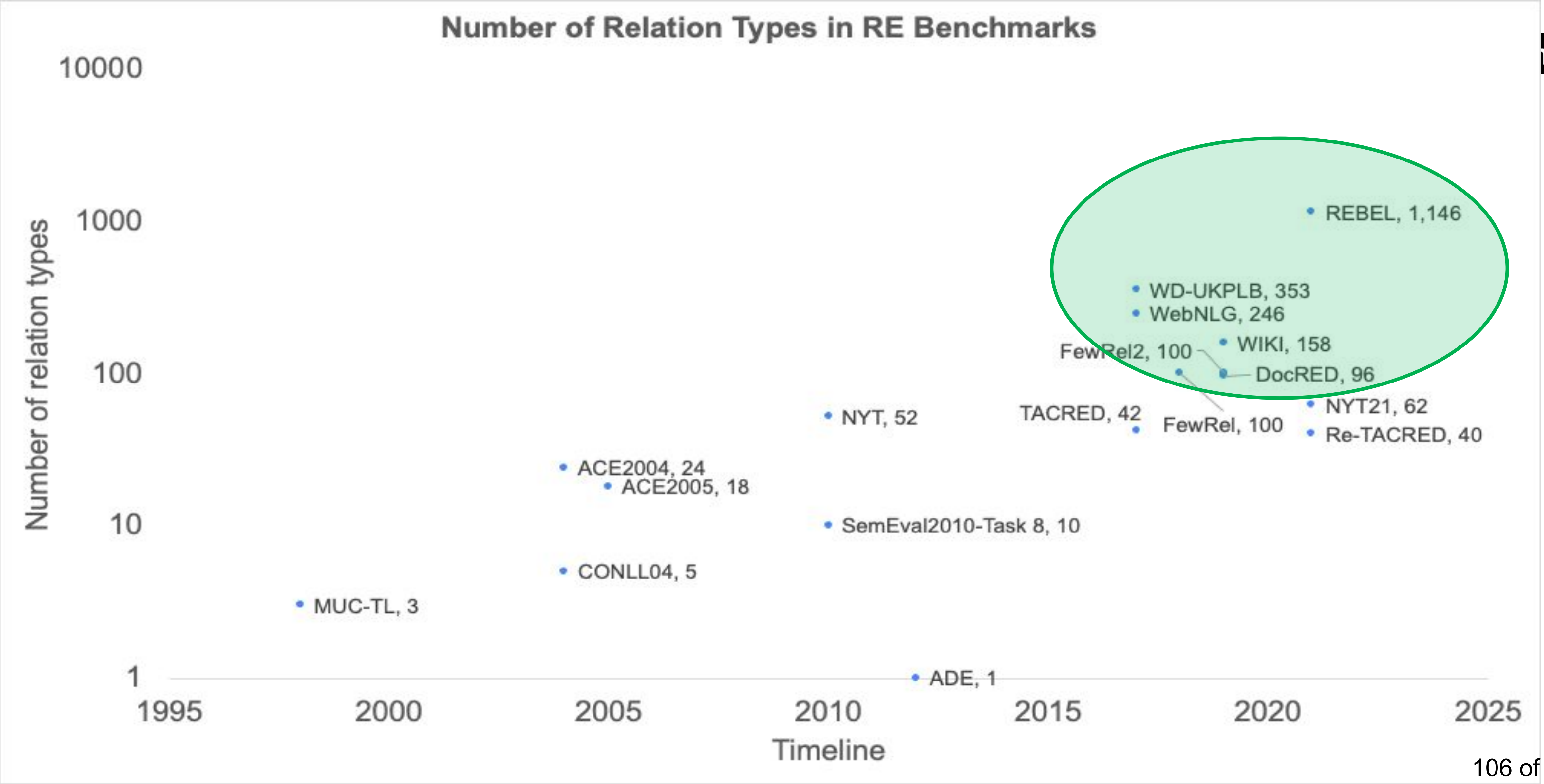


[DWIE] DWIE: An entity-centric dataset for multi-task document-level information extraction. Klim Zaporojets, Johannes Deleu, Chris Develder, Thomas Demeester. Information Processing and Management 2021.

IBM Research AI

Relation Extraction Academic Benchmarks

- 1998: MUC-TR (1998)
- 2004: CoNLL (2004); ACE (2004)
- 2005: ACE (2005)
- 2007: SemEval-07 Task 4 (2007)
- 2010: NYT (2010); SemEval2010 Task 8 (2010)
- 2012: ADE (2012)
- 2013: Google RE (2013)
- 2017: WebNLG (2017); TACRED (2017); WD-UKPLAB (2017); WebNLG (2017)
- 2018: FewRel (2018); T-REx (2018)
- 2019: FewRel2 (2019); DocRED (2019); WIKI (2019)
- 2021: NYT21 (2021); Re-TACRED (2021); DWIE (2021); REBEL (2021); KELM (2021); TEKGEN (2021)
- 2022: KERED (2022)



Number of Relation Types in RE Benchmarks
10000
1000
100
10
1
Number of relation types
1995
2000
2005
2010
2015
2020
2025
Timeline
REBEL, 1,146
WD-UKPLB, 353
WebNLG, 246
WIKI, 158
FewRel2, 100
DocRED, 96
NYT, 52
TACRED, 42
FewRel, 100
NYT21, 62
Re-TACRED, 40
ACE2004, 24
ACE2005, 18
SemEval2010-Task 8, 10
CONLL04, 5
MUC-TL, 3
ADE, 1

TIB

IBM Research AI

Domain-specific Relation Extraction: Bio Medical Domain

MEDLINE (2005)

i2b2 (2010)

Adverse Drug Events (ADE) (2012)

Drug-Drug Interaction (DDI) (2013)

BioNLP 2016 Shared Task

Chemical Disease Relation (CDR) (2016)

ChemProt (2016)

BioNLP 2019 Shared Task

Gene-Disease Associations GDA (2019)

CHemical Reactions (CHR) 2019

2005

2010

2012

2013

2016

2019

Approaches for Relation Extraction

IBM Research AI

Rule-based approaches
- Lexical analysis / phrase patterns
- Syntactical analysis and dependency Trees
- Semantic Parsing

Weak supervision / Distant supervision approaches
- Noise Reduction Approaches
- Embedding based approaches
- Auxiliary Information

Supervised approaches
- Feature based methods
- Kernel based methods
- Generative approaches (Seq2Seq)
- Word embeddings Based classifiers
- LSTMs, CNN, GNN
- Pretrained language model-based classifiers

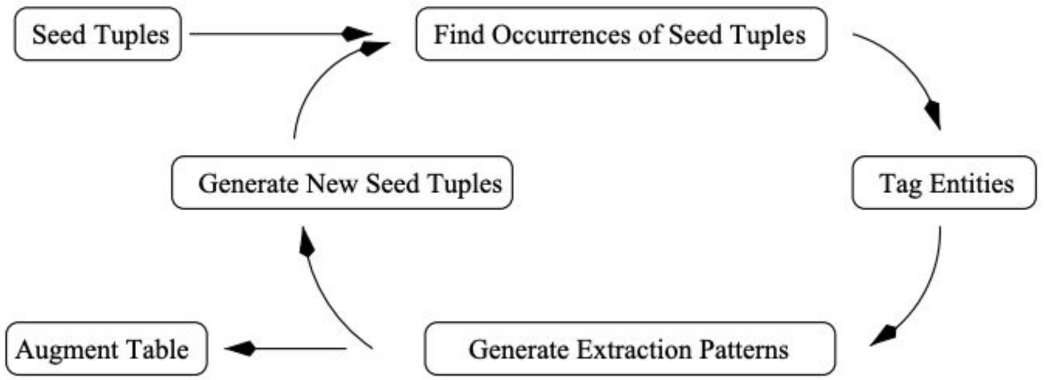
Joint Entity and Relation Extraction

A Review of Relation Extraction. Bach, Nguyen, and Sameer Badaskar. Literature Review for Language and Statistics (2007)
Relation Extraction using Distant Supervision: A Survey. Smirnova, A. and Cudré-Mauroux. ACM Computing Surveys (2018).
https://paperswithcode.com/task/relation-extraction
http://nlpprogress.com/english/relationship_extraction.html
https://github.com/roomylee/awesome-relation-extraction

Rule / Pattern-based approaches

IBM Research AI

headquarters relation



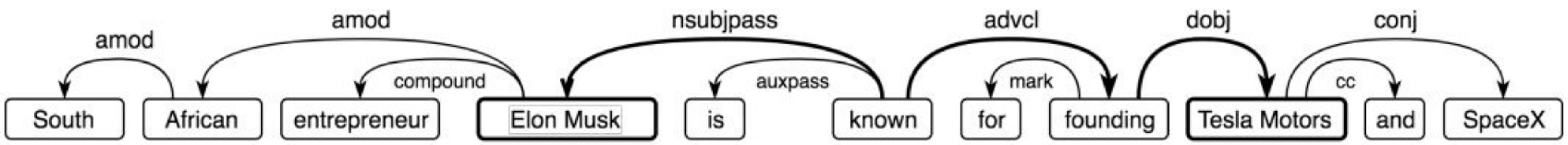
The main components of *Snowball*.

```
<ORGANIZATION>'s headquarters in <LOCATION>
<LOCATION>-based <ORGANIZATION>
<ORGANIZATION>, <LOCATION>
```

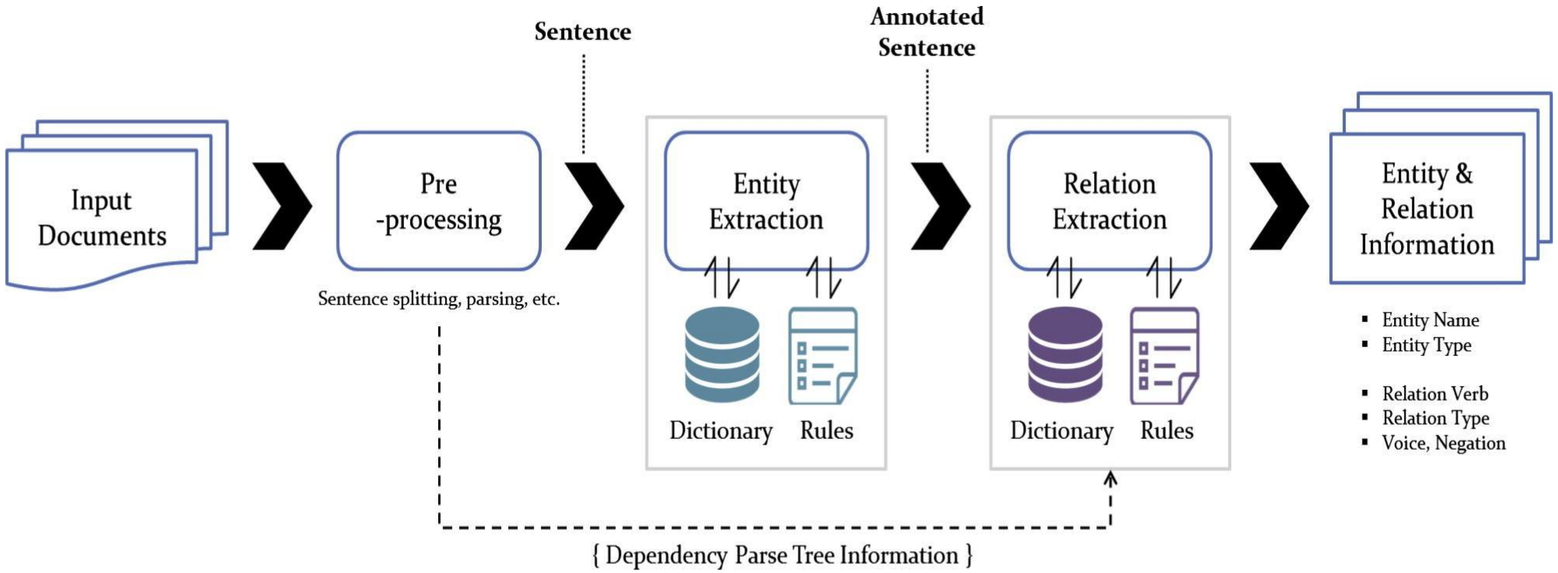
Snowball: Extracting Relations from Large Plain-Text Collections. Eugene Agichtein and Luis Gravano. DL2000.

Rule / Pattern-based approaches

IBM Research AI



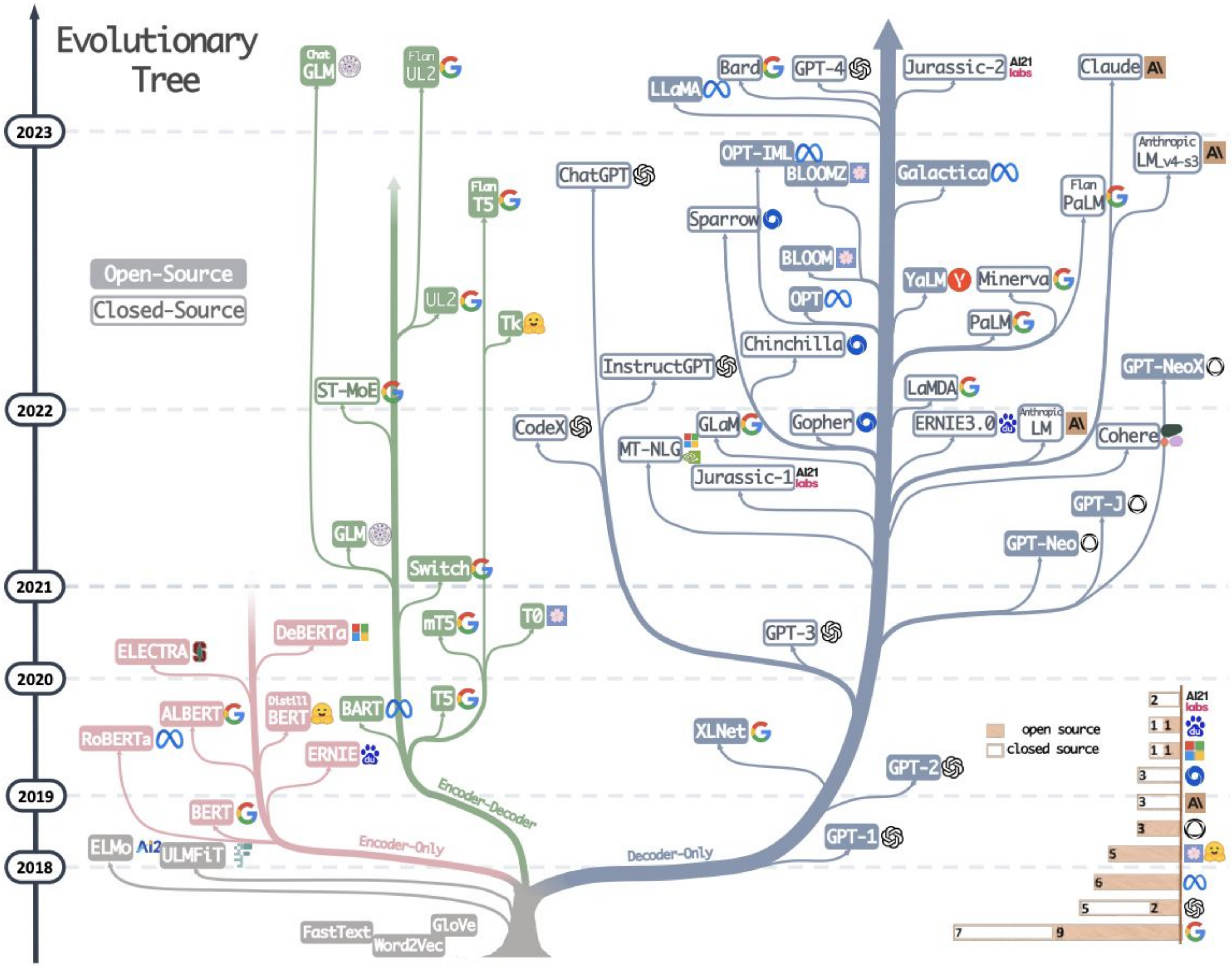
Dependency parsing



PKDE4J: Entity and relation extraction for public knowledge discovery. M Song, WC Kim, D Lee, GE Heo, KY Kang. Journal of Biomedical Informatics 2015.

IBM Research AI

Large Language Models / Foundation Models



Yang, Jingfeng, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, and Xia Hu. **"Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond."** https://arxiv.org/pdf/2304.13712.pdf

Relation extraction by classification (Encoder-Only)

IBM Research AI

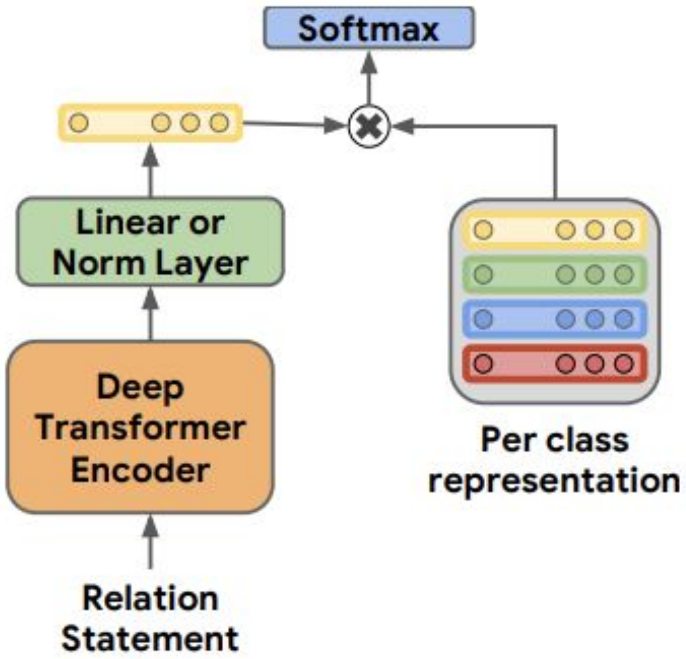


Fig. 2 in [Soares 2019]

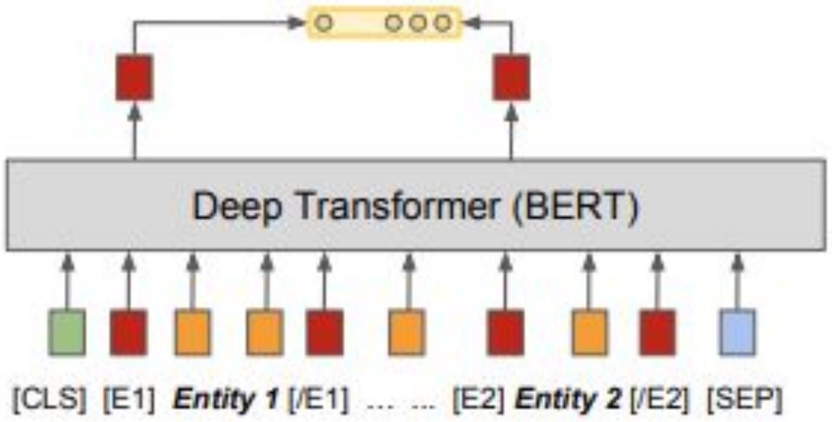


Fig. 3 in [Soares 2019]

[Soares 2019] Matching the Blanks: Distributional Similarity for Relation Learning. Livio Baldini Soares, Nicholas FitzGerald, Jeffrey Ling, Tom Kwiatkowski. ACL 2019.

Relation extraction with Seq2Seq models (Encoder-Decoder)

IBM Research AI

Irene Morgan, who was born and raised in Baltimore, lived on Long Island.

(Irene Morgan, birthPlace, Baltimore)
(Irene Morgan, residence, Long Island)

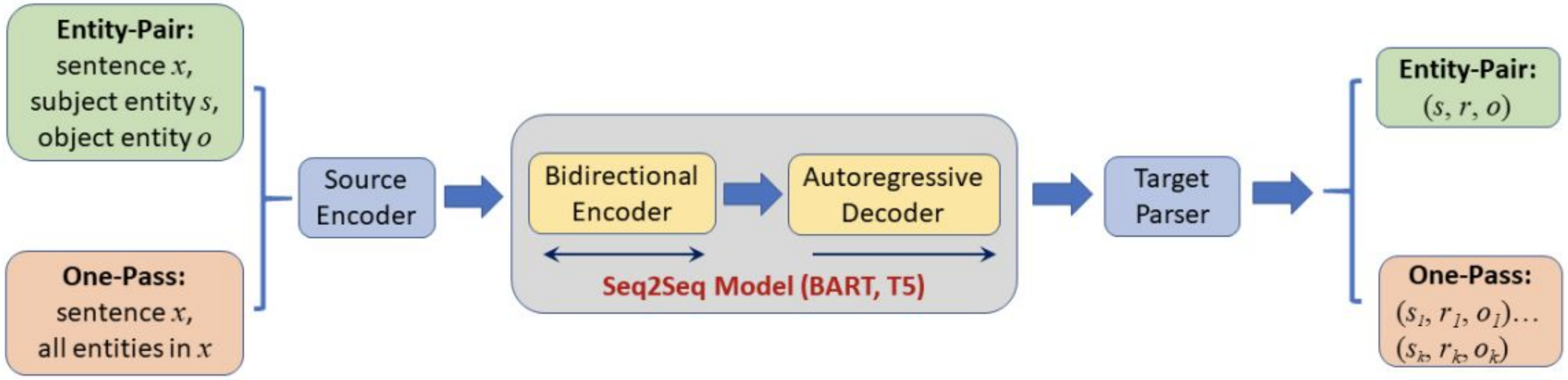
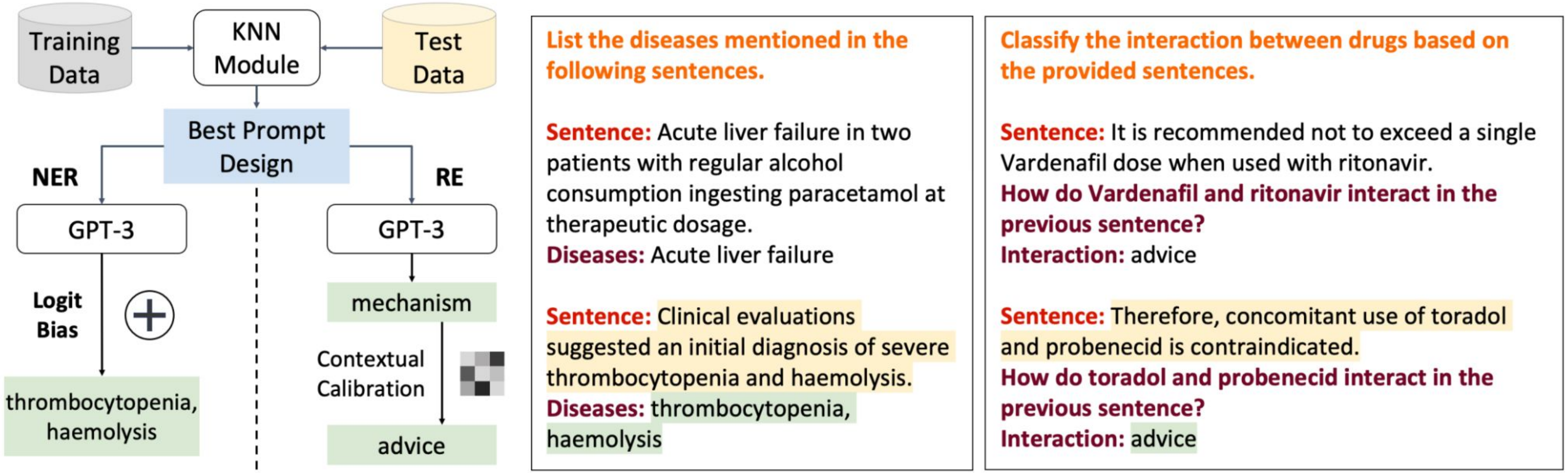


Fig 1 of [GREC]

- **[GenIE] GenIE: Generative Information Extraction**. Martin Josifoski, Nicola De Cao, Maxime Peyrard, Fabio Petroni, Robert West. NAACL 2022.
- **[REBEL] REBEL: Relation Extraction By End-to-end Language generation**. Pere-Lluís Huguet Cabot, Roberto Navigli. EMNLP 2021.
- **[GREC] A Generative Model for Relation Extraction and Classification**. Jian Ni, Gaetano Rossiello, Alfio Gliozzo, Radu Florian. Arxiv 2022.
- **[GenRL] Generative Relation Linking for Question Answering over Knowledge Bases**. Gaetano Rossiello, Nandana Mihindukulasooriya, Ibrahim Abdelaziz, Mihaela Bornea, Alfio Gliozzo, Tahira Naseem, Pavan Kapanipathi. ISWC 2021.

Relation Extraction with In-Context Learning (Decoder-Only)

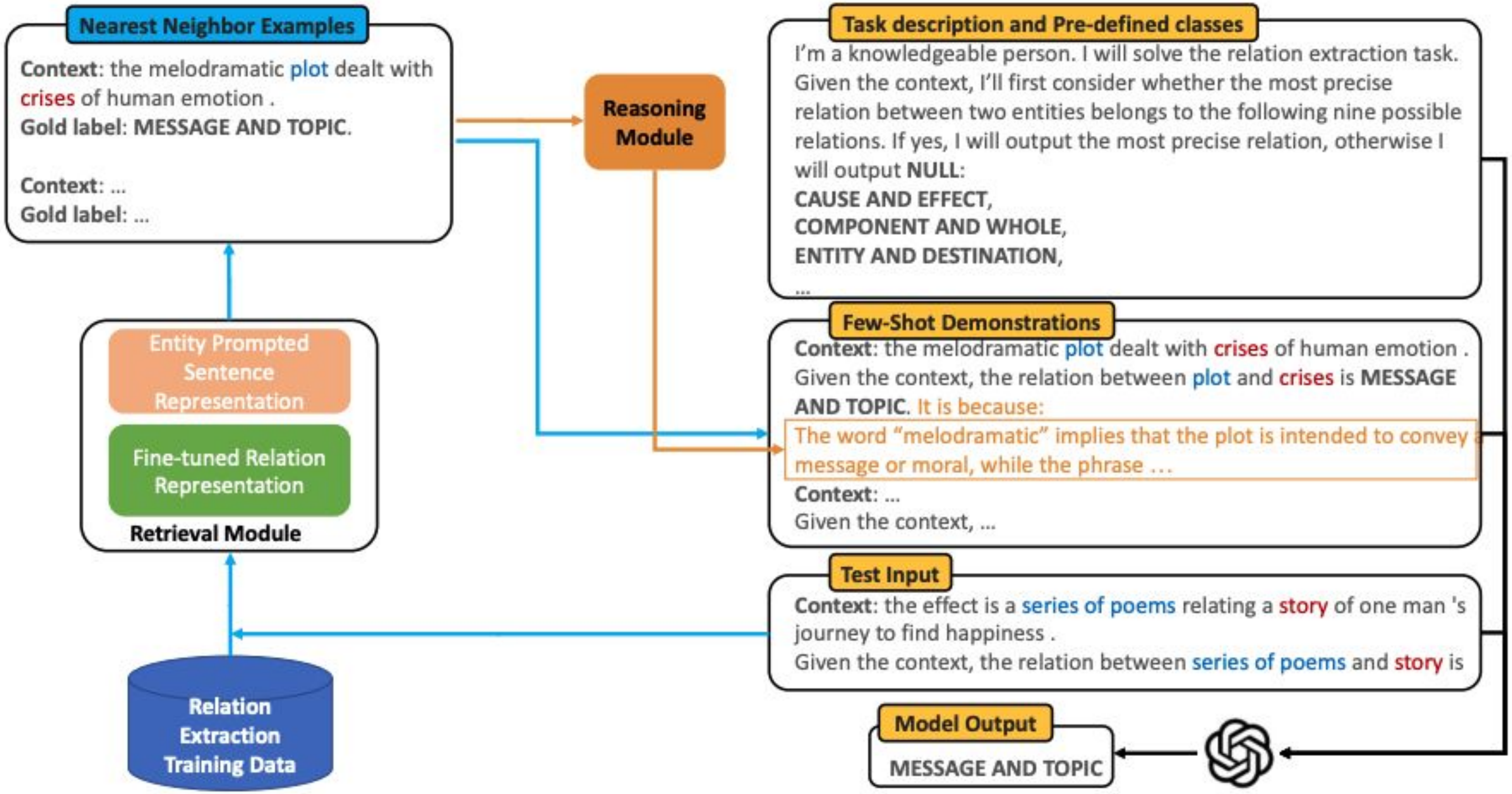
IBM Research AI



[Gutierrez2022] **Thinking about GPT-3 In-Context Learning for Biomedical IE? Think Again** (Jimenez Gutierrez et al., Findings 2022)

Relation Extraction with In-Context Learning

IBM Research AI



[GPTRE2023] GPT-RE: In-context Learning for Relation Extraction using Large Language Models.
Zhen Wan, Fei Cheng, Zhuoyuan Mao, Qianying Liu, Haiyue Song, Jiwei Li, Sadao Kurohashi.

Recommended resources

Natural Language Processing

Relation Extraction

Edit

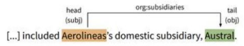
416 papers with code • 40 benchmarks • 52 datasets

Benchmarks

Add a Result

These leaderboards are used to track progress in Relation Extraction

Trend	Dataset	Best Model	Paper	Code	Compare
	DocRED	KD-Rb-l			See all
	TACRED	RECENT+SpanBERT			See all
	ACE 2005	PL-Marker			See all
	SemEval-2010 Task 8	QA			See all
	NYT	REBEL			See all
	CoNLL04	REBEL			See all
	Adverse Drug Events (ADE) Corpus	Spark NLP			See all
	WebNLG	PFN			See all



Content

- Introduction
- Benchmarks
- Datasets
- Subtasks
- Libraries
- Papers
  - Most implemented
  - Social
  - Latest
  - No code

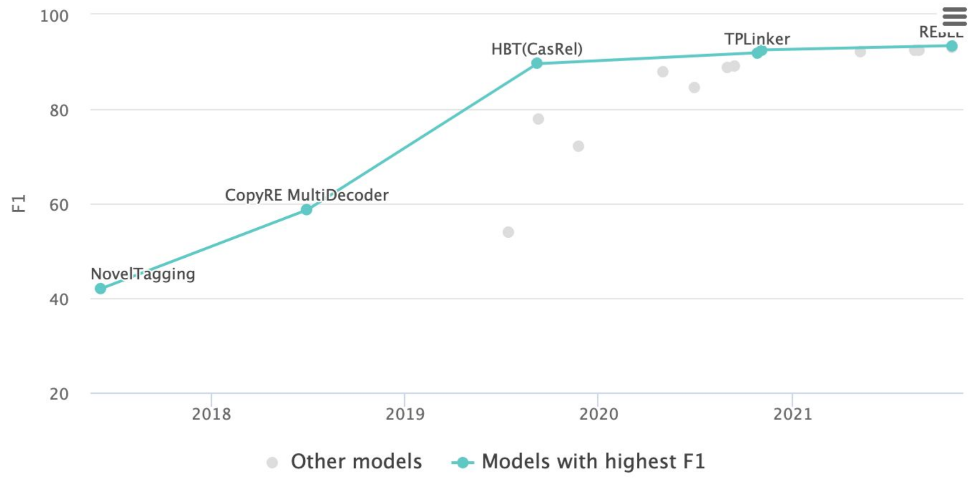
Relation Extraction

Relation Extraction on NYT

Leaderboard | Community Models | Dataset

View F1 by Date for All models

All competition entries



https://paperswithcode.com/task/relation-extraction

Some take home messages

IBM Research AI

- There is a growing interest in the field of relation extraction and applying that to building large scale knowledge graphs

- More complex and realistic relation linking benchmarks are being proposed with larger number of relations, document-level context, etc.

- Recent advancements in NLP with transformer-based pretrained language models and generative approaches pushing the state of the art

- Advancements in relation extraction can help both academic and industry move towards automatically building knowledge graphs from text.

References:

IBM Research AI

[FewRel] FewRel: A Large-Scale Supervised Few-Shot Relation Classification Dataset with State-of-the-Art Evaluation. Xu Han, Hao Zhu, Pengfei Yu, Ziyun Wang, Yuan Yao, Zhiyuan Liu, Maosong Sun. EMNLP2018.

[FewRel2] FewRel 2.0: Towards More Challenging Few-Shot Relation Classification. Tianyu Gao, Xu Han, Hao Zhu, Zhiyuan Liu, Peng Li, Maosong Sun, Jie Zhou. EMNLP2019.

[SemEval-2010 Task 8] SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals. Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, Stan Szpakowicz. SemEval - ACL 2010.

[NYT] Modeling Relations and Their Mentions without Labeled Text. Sebastian Riedel, Limin Yao & Andrew McCallum. ECML PKDD 2010.

[TACRED] Position-aware Attention and Supervised Data Improve Slot Filling. Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, Christopher D. Manning. EMNLP 2017.

[ACE 2004] ACE 2004 Multilingual Training Corpus. Alexis Mitchell, Stephanie Strassel, Shudong Huang, Ramez Zakhary. LDC2005.

[ACE 2005] ACE 2005 Multilingual Training Corpus. Christopher Walker, Stephanie Strassel, Julie Medero, Kazuaki Maeda. LDC2006.

[DocRED] DocRED: A Large-Scale Document-Level Relation Extraction Dataset. Yuan Yao, Deming Ye, Peng Li, Xu Han, Yankai Lin, Zheng-Hao Liu, Zhiyuan Liu, Lixin Huang, Jie zhou, Maosong Sun. ACL 2019.

References:

IBM Research AI

[CoNLL04] A Linear Programming Formulation for Global Inference in Natural Language Tasks. Dan Roth, Wen-tau Yih. NAACL 2004.

[Re-TACRED] Re-TACRED: Addressing Shortcomings of the TACRED Dataset. George Stoica, Emmanouil Antonios Platanios, Barnabás Póczos. AAAI 2021.

[ADE] Development of a Benchmark Corpus to Support the Automatic Extraction of Drug-related Adverse Effects from Medical Case Reports. Journal of Biomedical Informatics 2012.

[WebNLG] Creating Training Corpora for NLG Micro-Planners. Claire Gardent, Anastasia Shimorina, Shashi Narayan, Laura Perez-Beltrachini. ACL2017.

[NYT21] Revisiting the Negative Data of Distantly Supervised Relation Extraction. Chenhao Xie, Jiaqing Liang, Jingping Liu, Chengsong Huang, Wenhao Huang, Yanghua Xiao. ACL 2021.

[DWIE] DWIE: An entity-centric dataset for multi-task document-level information extraction. Klim Zaporojets, Johannes Deleu, Chris Develder, Thomas Demeester. Information Processing and Management 2021.

[GIDS] Improving Distantly Supervised Relation Extraction using Word and Entity Based Attention. Sharmistha Jat, Siddhesh Khandelwal, Partha Talukdar. AKBC 2017.

[WIKI] Neural Relation Extraction for Knowledge Base Enrichment. Bayu Distiawan Trisedya, Gerhard Weikum, Jianzhong Qi, Rui Zhang. ACL2019.

References:

IBM Research AI

[GDA] RENET: A Deep Learning Approach for Extracting Gene-Disease Associations from Literature. Ye Wu, Ruibang Luo, Henry C. M. Leung, Hing-Fung Ting & Tak-Wah Lam. RECOMB 2019.

[ChemProt] Martin Krallinger et al. Overview of the BioCreative VI chemical-protein interaction track. BioCreative 2017.

[CDR] BioCreative V CDR task corpus: a resource for chemical disease relation extraction. Jiao Li et al. Database : the journal of biological databases
and curation. 2016.

[DDI] The DDI corpus: an annotated corpus with pharmacological substances and drug-drug interactions. Maria Herrero-Zazo, Isabel Segura-Bedmar, Paloma Martinez, and Thierry Declerck. Journal of biomedical informatics 2013.

[ADE] Development of a Benchmark Corpus to Support the Automatic Extraction of Drug-related Adverse Effects from Medical Case Reports. Journal of Biomedical Informatics 2012.

[i2b2 2010] Ozlem Uzuner, Brett R South, Shuying Shen, and Scott DuVall. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. Journal of the American Medical Informatics Association 2011.

[MEDLINE] A shortest path dependency kernel for relation extraction. Bunescu, R. C., & Mooney, R. J. HLT ’2005.