Utilising of Terminological Expressions in Knowledge Bases and Knowledge Graphs for Neural Machine Translation

Mihael Arcan
Unit for Natural Language Processing
Insight SFI Research Centre for Data Analytics
Data Science Institute, NUI Galway
mihael.arcan@nuigalway.ie

based on:

Translating Terminological Expressions in Knowledge Bases with Neural Machine Translation

Mihael Arčan, Daniel Torregrosa and Paul Buitelaar
Insight Centre for Data Analytics, Data Science Institute
National University of Ireland Galway
`[firstname.lastname]@insight-centre.org`

Abstract

Our work presented in this paper focuses on the translation of terminological expressions represented in semantically structured resources, like ontologies or knowledge graphs. The challenge of translating ontology labels or terminological expressions documented in knowledge bases lies in the highly specific vocabulary and the lack of contextual information, which can guide a machine translation system to translate ambiguous words into the targeted domain. Due to these challenges, we evaluate the translation quality of domain-specific expressions in the medical and financial domain with statistical as well as with neural machine translation methods and experiment domain adaptation of the translation models with terminological expressions only. Furthermore, we perform experiments on the injection of external terminological expressions into the translation systems. Through these experiments, we observed a significant advantage in domain adaptation for the domain-specific resource in the medical and financial domain and the benefit of subword models over word-based neural machine translation models for terminology translation.

1 Introduction

Most of the labels stored in semantically struc-

2013; Arcan et al., 2013), providing information related to an ontology label, e.g. *other intangible assets*, in Spanish, German or Italian.

Due to the large success of neural machine translation (NMT) (Kalchbrenner and Blunsom, 2013; Sutskever et al., 2014), we compare the use of NMT and statistical machine translation (SMT) (Brown et al., 1993) when translating terminological expressions in isolation, i.e. when they do not form part of a whole sentence. This is motivated by the invasive manual procedure domain experts would alternatively endure. Although automatically generated translations of these domain-specific expressions are far from perfect, studies have shown significant productivity gains when human translators are supported by machine translation output rather than starting a translation task from scratch (Federico et al., 2012; Läubli et al., 2013; Green et al., 2013).

For both translation methods, we translated the ontology labels in the medical and financial domain, documented in the International Classification of Diseases (ICD) and in the International Financial Reporting System (IFRS) ontology. Furthermore, we translated the Wikipedia titles, which represent a mixture of generic as well as domain-specific expressions. Since large paral-

Utilizing Knowledge Graphs for Neural Machine Translation Augmentation

Diego Moussallem
Data Science Research Group, Paderborn University
Germany
diego.moussallem@upb.de

Axel-Cyrille Ngonga Ngomo
Data Science Group, Paderborn University
Germany
axel.ngonga@upb.de

Paul Buitelaar
Insight Centre for Data Analytics, Data Science Institute
National University of Ireland Galway
Ireland
paul.buitelaar@insight-centre.org

Mihael Arcan
Insight Centre for Data Analytics, Data Science Institute
National University of Ireland Galway
Ireland
mihael.arcan@insight-centre.org

ABSTRACT

While neural networks have led to substantial progress in machine translation, their success depends heavily on large amounts of training data. However, parallel training corpora are not always readily available. Moreover, out-of-vocabulary words—mostly entities and terminological expressions—pose a difficult challenge to Neural Machine Translation systems. Recent efforts have tried to alleviate the data sparsity problem by augmenting the training data using different strategies, such as external knowledge injection. In this paper, we hypothesize that knowledge graphs enhance the semantic feature extraction of neural models, thus optimizing the translation of entities and terminological expressions in texts and consequently leading to better translation quality. We investigate two different strategies for incorporating knowledge graphs into neural models without modifying the neural network architectures. Additionally, we examine the effectiveness of our augmented models on domain-specific texts and ontologies. Our knowledge-graph-augmented neural translation model, dubbed *KG-NMT*, achieves significant and consistent improvements of +3 BLEU, METEOR and CHRF3 on average on the *newstest* datasets between 2015 and 2018 for the WMT English-German translation task.

CCS CONCEPTS

• Information systems → Information extraction; • Computing methodologies → Natural language processing.

KEYWORDS

Neural Machine Translation, Knowledge Graphs, NLP, Linked Data

ACM Reference Format:

Diego Moussallem, Axel-Cyrille Ngonga Ngomo, Paul Buitelaar, and Mihael Arcan. 2019. Utilizing Knowledge Graphs for Neural Machine Translation Augmentation. In *Proceedings of the 10th International Conference on Knowledge Capture (K-CAP '19), November 19–21, 2019, Marina Del Rey, CA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3360901.3364423

1 INTRODUCTION

Neural Network (NN) models have shown significant improvements in translation generation and have been widely adopted due to their consistent improvements over the Phrase-Based Statistical Machine Translation (PBSMT) approaches [20]. A number of Neural Machine Translation (NMT) architectures have been proposed in recent years, ranging from recurrent [4] to self-attentional networks [45]. However, a major drawback of these models is that they need large amounts of training data to return adequate results and have a limited vocabulary size due to their computational complexity [25]. The data sparsity problem in Machine Translation (MT), which is mostly caused by the lack of training data, manifests itself in particular in a poor translation of out-of-vocabulary (OOV) words, e.g., entities or terminological expressions rarely or never seen in the training data.

Previous work has attempted to deal with the data sparsity problem by introducing character-based models [25] and Byte Pair Encoding (BPE) algorithms [38]. Also, different strategies were devised for overcoming the lack of training data, for instance, back-translation [37], which relies on the use of monolingual data being translated by a different NMT model and added as additional synthetic training data. In addition, pre-trained monolingual word embeddings were used for leveraging the embedding matrix of low-resource NMT models [34]. Moreover, distinct external knowledge injection approaches have been tried for guiding the NNs to use domain-specific data while translating, for example, ontologies [1, 2].

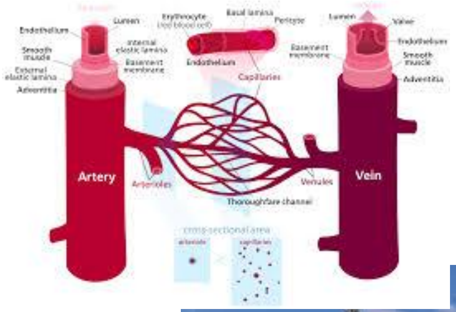
Despite the significant advancement of previous work in NMT, translating entities and terminological expressions remains a chal-

NUI Galway
OÉ Gaillimh

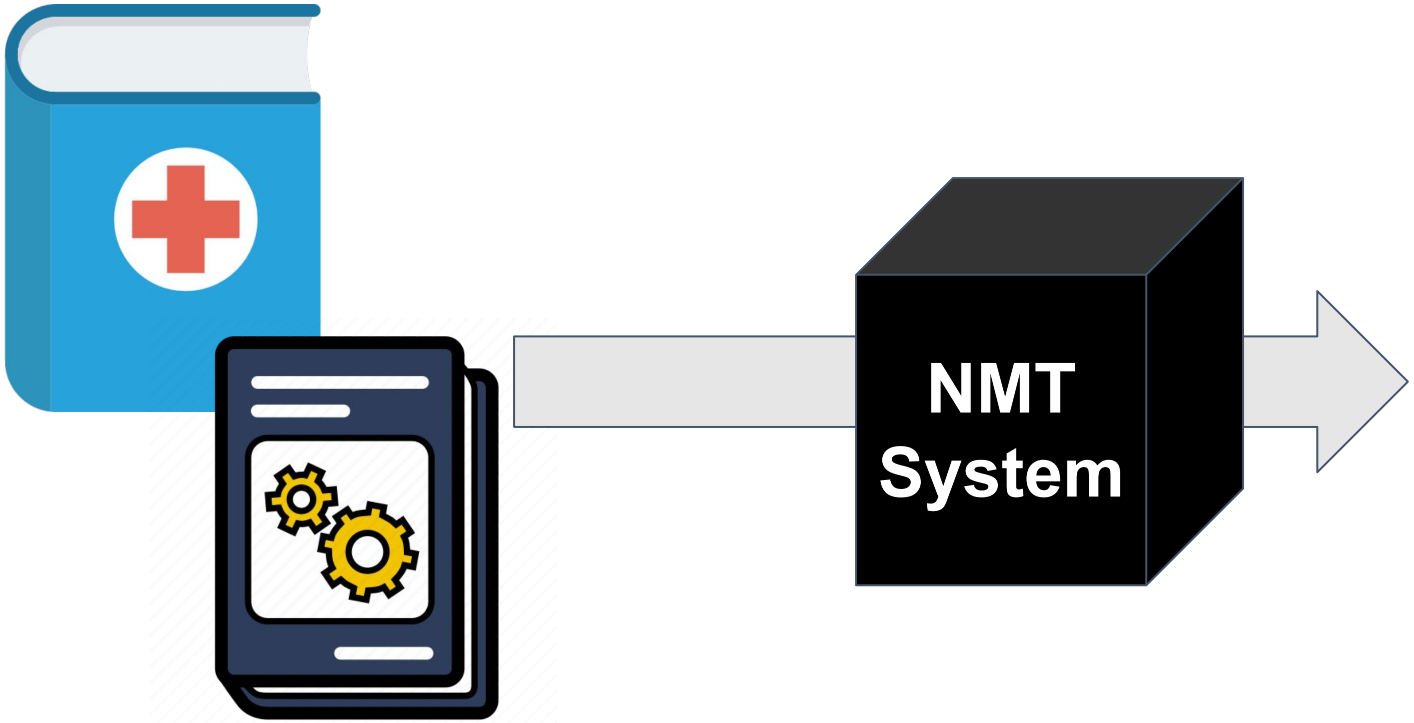
Open Issues in Machine Translation

- Named Entities, Terminological expressions
  - semantic ambiguity
    (Paris, Smith, … , vessel, injection, hedges …)

  - data sparsity, i.e. lack of training data
    (*staycation, MLT*)
    - Out-Of-Vocabulary Words

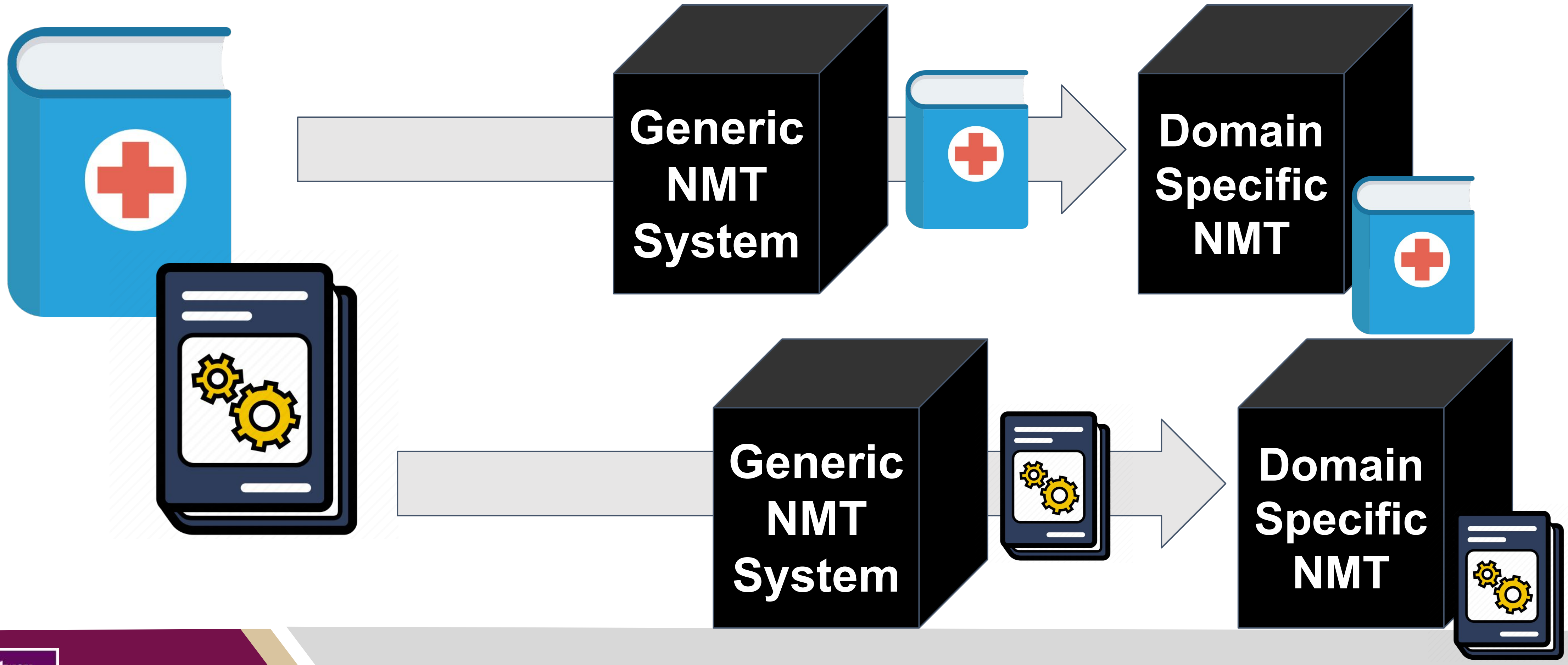


Domains?



- Do we translate text with a generic model?
- Do we re-train a new translation system for each domain?
- What is the best domain adaptation approach?

Domain Adaptation with Continuous Training



Terminology Translation

E.g. ICD Terms:

- Intestinal infectious diseases
- Infectious gastroenteritis and colitis, unspecified
- Viral and other specified intestinal infections
- Other protozoal intestinal diseases
- Amebiasis
- Other bacterial foodborne intoxications, not elsewhere classified
- Other bacterial intestinal infections
- Shigellosis
- Other salmonella infections
- Typhoid and paratyphoid fevers

Terminology Translation

E.g. IFRS Terms:

- Vehicles, expenditures recognised for constructions
- Vehicles, gross
- Vehicles, revaluation surplus
- Wages and salaries
- Warranty provision
- Weighted average exercise price of share options exercised, share-based payment arrangement
- Weighted average shares
- Weighted average shares and adjusted weighted average shares

Terminology Translation Evaluation

	BLEU			METEOR			chrF3		
NMT models	ICD_{eval}	$IFRS_{eval}$	$Wiki_{eval}$	ICD_{eval}	$IFRS_{eval}$	$Wiki_{eval}$	ICD_{eval}	$IFRS_{eval}$	$Wiki_{eval}$
Baseline	3.20	9.35	8.20	8.65	13.18	10.13	20.46	28.35	21.38
ICD_{dev}	**20.89**	3.55	3.21	**31.06**	12.14	6.86	**37.00**	15.08	11.06
$IFRS_{dev}$	2.53	**58.17**	7.39	13.62	**65.48**	14.71	20.90	**65.47**	18.05
$Wiki_{dev}$	1.00	0.74	**26.27**	3.99	5.27	**27.08**	7.51	8.32	**26.64**
NMT_{BPE} models	ICD_{eval}	$IFRS_{eval}$	$Wiki_{eval}$	ICD_{eval}	$IFRS_{eval}$	$Wiki_{eval}$	ICD_{eval}	$IFRS_{eval}$	$Wiki_{eval}$
Baseline	4.29	13.55	13.51	19.22	30.31	26.90	38.99	43.48	45.01
ICD_{dev}	**50.15**	5.76	9.41	**58.68**	19.05	20.71	**72.82**	35.05	38.29
$IFRS_{dev}$	4.86	**75.03**	10.78	20.83	**81.48**	23.90	40.70	**88.22**	42.35
$Wiki_{dev}$	1.53	1.52	**41.19**	7.69	8.98	**42.34**	22.47	19.35	**55.76**

Term Identification and Disambiguation

*Dass deine **Stammzellen** im **KM** trotz **Bestrahlung** so gut drauf sind und dir **Leukos** bauen, freut mich total! Zu den ….*

Stammzelle

(Weitergeleitet von Stammzellen)

Als **Stammzellen** werden allgemein Körperzellen bezeichnet, die sich in verschiedene Zelltypen oder Gewebe ausdifferenzieren können. Je nach Art der Stam
haben sie das Potenzial, sich in jegliches Gewebe (embryonale Stammzellen) oder in bestimmte festgelegte Gewebetypen (adulte Stammzellen) zu entwickeln

Stammzellen sind in der Lage, Tochterzellen zu generieren, die selbst wiederum Stammzelleigenschaften besitzen, aber auch solche mit größerer Ausdifferenz
noch nicht vollständig geklärter Mechanismus asymmetrischer Zellteilung. Über das jeweilige Schicksal der Zellen entscheidet dabei vor allem das biologische

Stammzellen werden vor allem durch ihr ontogenetisches Alter und ihr Differenzierungspotenzial unterschieden: die ontogenetisch frühesten Stammzellen sir
Stammzellen, aus denen später die primitiven Keimstammzellen sowie die somatischen Stamm- und Progenitorzellen (oder Vorläuferzellen) hervorgehen. Phy
Stammzellen auf den letzten gemeinsamen eukaryotischen Vorfahren (LECA) zurück.[2]

Auch Pflanzen besitzen Stammzellen. Diese befinden sich an der Spitze des Sprosses im sogenannten Apikalmeristem sowie an den Wurzelspitzen im Wurzeln
allen tierischen und menschlichen Zellen besitzen bei Pflanzen praktisch alle Zellen die Fähigkeit, einen kompletten Organismus zu regenerieren.

KM

KM steht als Abkürzung für:

- Air Malta, nach dem IATA-Code
- Kaltmiete, siehe Liste der Abkürzungen in Wohnungsanzeigen
- Kampfmannschaft, siehe Vereinsmannschaft
- Katholischer Medienverband
- Kindesmutter, Amtssprache für Mutter
- K&M Elektronik, ein süddeutsches Handelsunternehmen für Comp
- KM (Ekranoplan), großes sowjetisches Bodeneffektfahrzeug aus de
- KM-Stoffe, chemische Substanzen, die karzinogen und/oder mutag
- Knochenmark
- Knowledge Management, siehe Wissensmanagement
- Koleje Mazowieckie, polnisches Eisenbahnunternehmen
- Kommunikationsmanagement, siehe Öffentlichkeitsarbeit
- Komoren als Ländercode nach ISO 3166-2
- Komplementärmedizin, siehe Alternativmedizin
- Konfigurationsmanagement
- Koninklijke Marine, die niederländische Marine
- Kontrastmittel
- Kontrollmitteilung im Steuerrecht
- Konvertible Mark, eine Währung in Bosnien-Herzegowina
- Korrespondierendes Mitglied, siehe Mitglied
- Krause-Mishler, ein System um Münzen zu nummerieren, das von
- Kriminalmeister, ein Dienstgrad bei der deutschen Polizei, siehe Po

Using Wikipedia for Term Identification

Step 1 — Identify Terms within Documents

Step 2 — Disambiguate Identified Terms (Document vs Wiki Abstract similarity)

Step 3 — Continue Training the Generic NMT system

Step 1 — Identify Terms within Documents

Step 2 — Identify Wikipedia Categories associated with Identified Terms

Step 3 — Use all Wikipedia terms, with top-k Wikipedia Categories

Step 4 — Continue Training the Generic NMT system

Results on Term Translation

(dict.cc, MedTerm, Wiki)

Model	German → English			French → English			Chinese → English		
	BLEU-1	METEOR	chrF	BLEU-1	METEOR	chrF	BLEU-1	METEOR	chrF
generic model	6.7	7.7	42.8	48.9	30.2	57.0	2.2	8.5	21.5
+ cont. (Wikipedia Terms, + 0.1 SemSim)	7.7	12.7	42.9	52.2	30.9	59.8	**23.7**	**11.9**	**23.6**
+ cont. (WikiCategories/ Terms)	10.2	11.5	48.8	52.1	33.1	62.3	/	/	/
+ cont. (ICD)	14.4	12.2	55.0	/	/	/	/	/	/
+ cont. (dict.cc/MedTerm, bpe32k)	**25.5**	**34.2**	**61.3**	**56.6**	**36.7**	**64.3**	/	/	/

NUI Galway
OÉ Gaillimh

German/French → English on EMEA Dataset

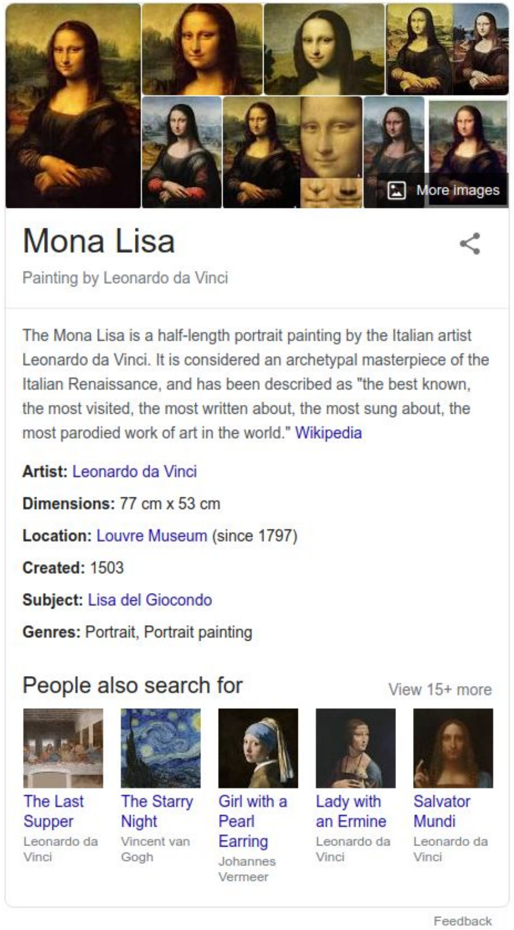
Model	German → English			French → English		
	BLEU	METEOR	chrF	BLEU	METEOR	chrF
generic model	14.29	20.63	36.59	21.36	24.95	40.59
+ cont. (Wiki Terms, 0.1 Semantic Similarity)	7.97	14.11	26.16	13.21	17.81	29.88

NUI Galway
OÉ Gaillimh

Knowledge Graph

- is a knowledge base of a domain knowledge, created by a domain experts
- core of many of the tools that we use in our daily lives, such as voice assistants (Alexa, Siri or Google Assistant)

Google KG



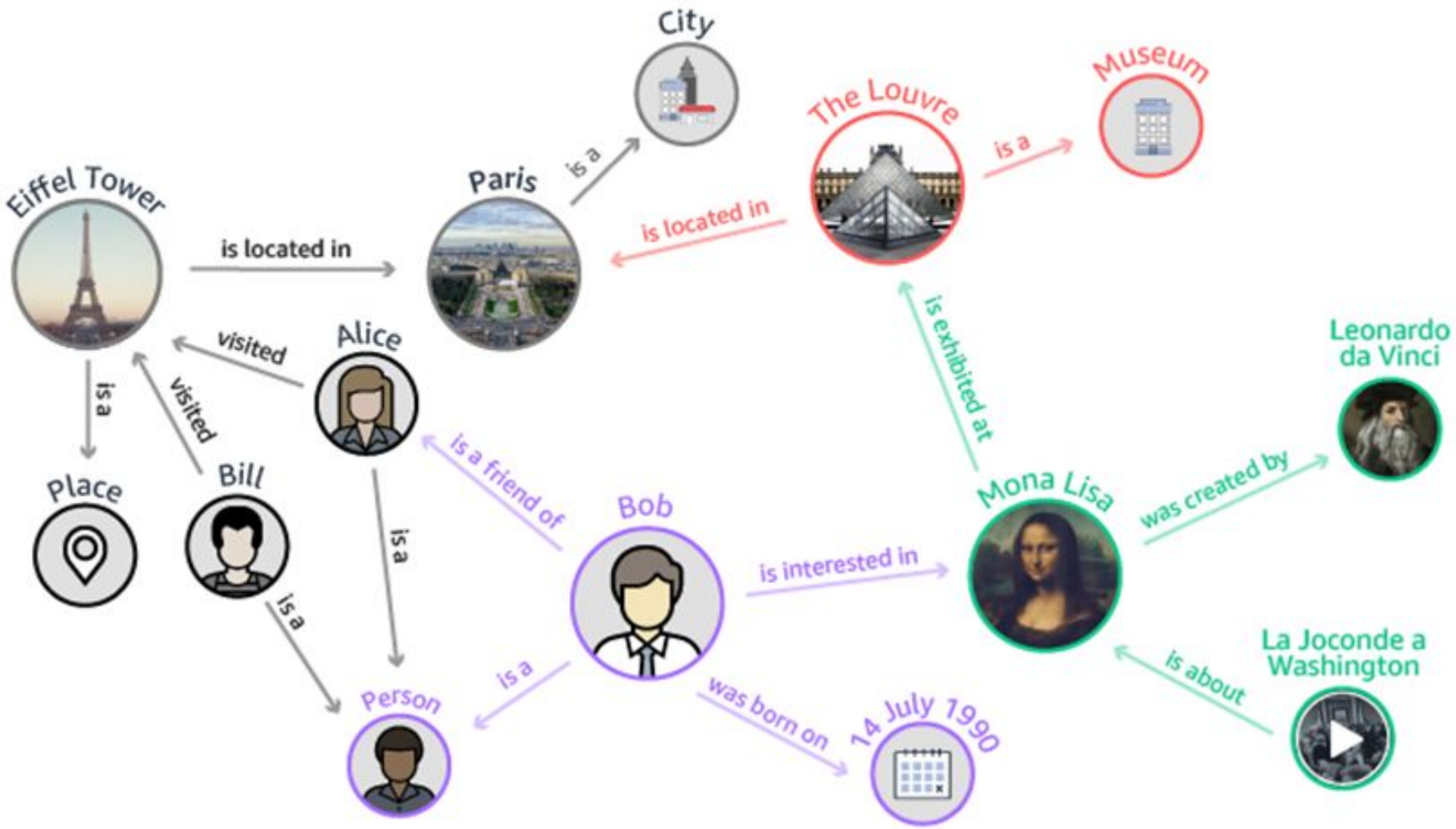
Wikipedia Infobox

Addressing "Translation Issues with Entities"

- Question: How can we improve the translation quality of entities and alleviate the Out-Of-Vocabulary (OOV) issue?

- Answer:
  - Knowledge Graph-augmented NMT (KG-NMT)

Why Knowledge Graphs?



https://aws.amazon.com/blogs/apn/exploring-knowledge-graphs-on-amazon-neptune-using-metaphactory/

KG Represenations

- RDF triples (Subject, Predicate, Object)

…

<.../Mona_Lisa> <.../property/title> "Mona Lisa"@en

<.../Mona_Lisa> <.../property/artist> <.../Leonardo Da Vinci >

<.../Mona_Lisa> <.../property/museum> <.../Louvre>

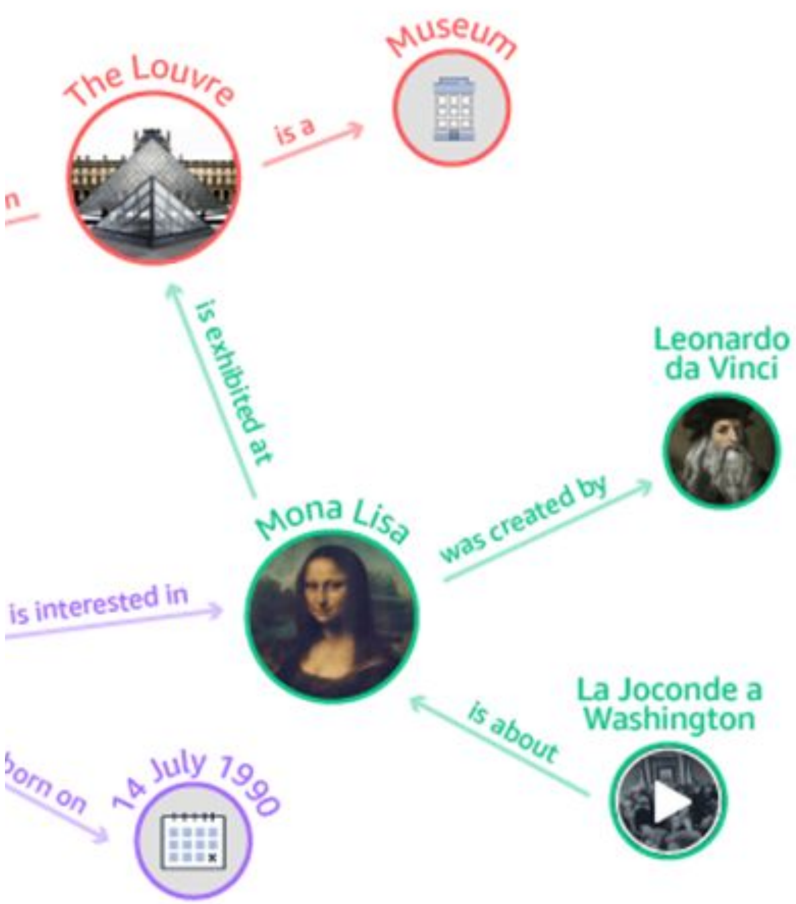
…

<.../Louvre> </property/name> "The Louvre Museum"@en .

<.../Louvre> <…/property/type> "Art museum, Design/Textile Museum, Historic site"@en .

<.../Louvre> <…/property/director> "Jean-Luc Martinez"@en

<.../Louvre> <…/property/location> "France"@en

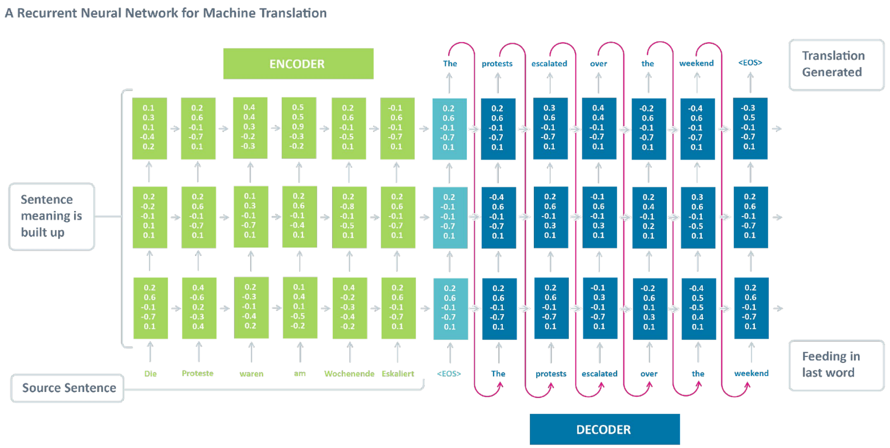


RDF – Resource Description Framework

NUI Galway
OÉ Gaillimh

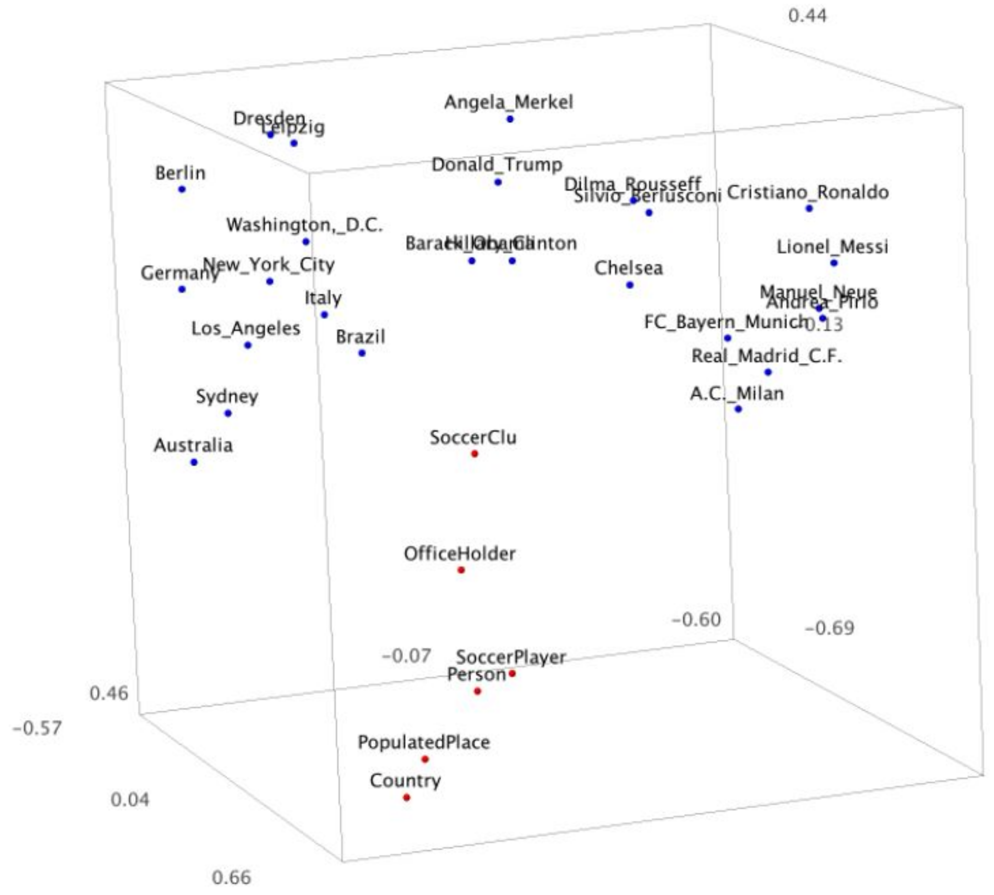
Incorporating RDF triples into MT

…
<.../Mona_Lisa> <.../property/title> "Mona Lisa"@en .
<.../Mona_Lisa> <.../property/museum> <.../Louvre>
…
<.../Louvre> <…/property/location> "France"@en
<…/France> <…/owl#sameAs> “Frankreich”@de
…



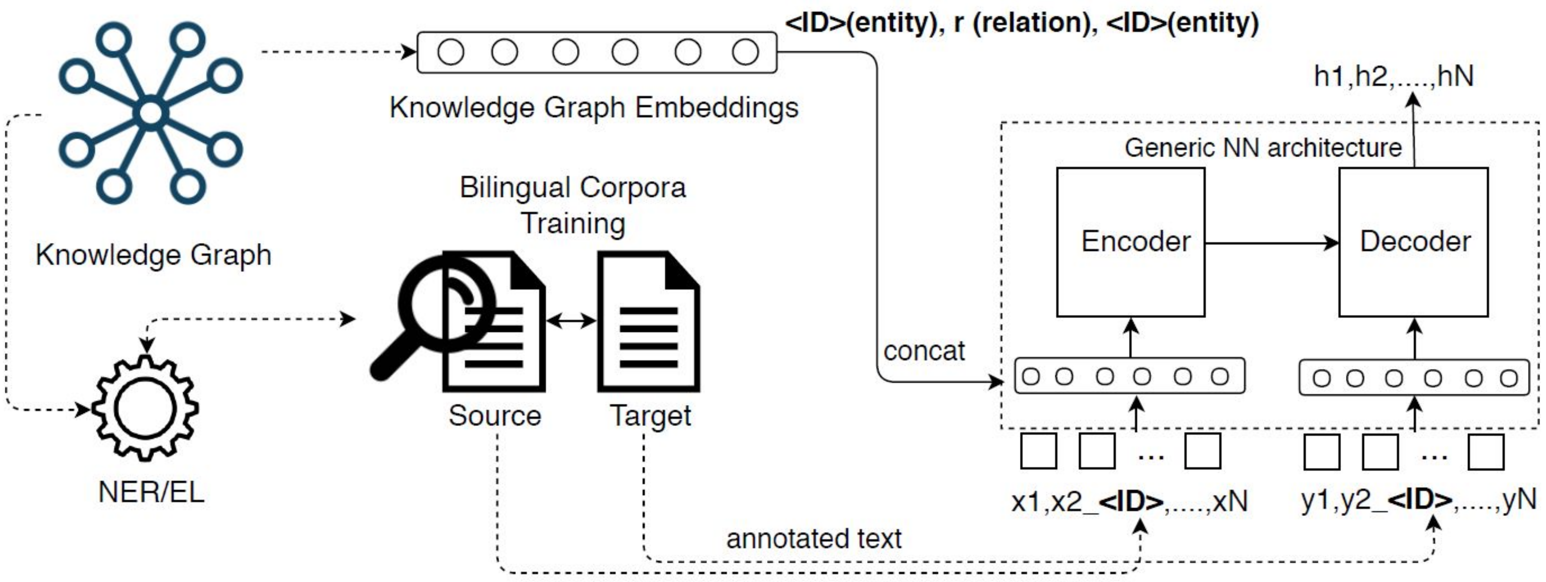
http://naviglinlp.blogspot.com/2018/06/lecture-26-neural-machine-translation.html

Knowledge Graph Embeddings (KGE)



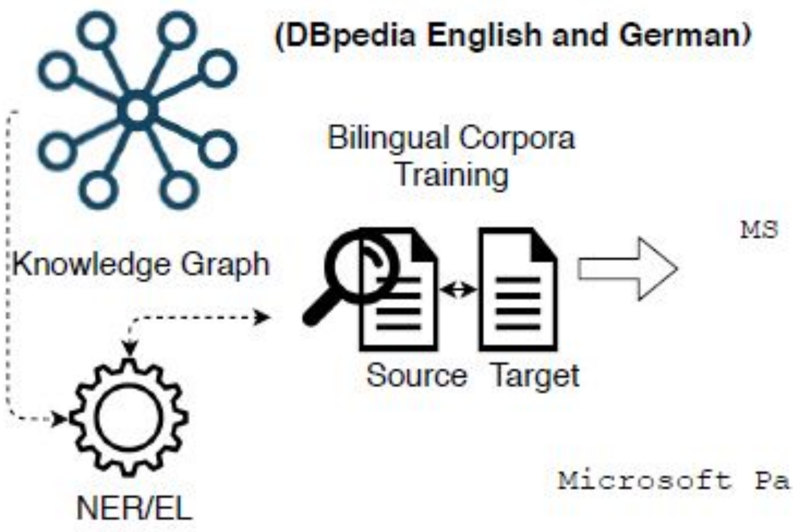
Proposed KG-NMT Methodology

- Entity Linking[1] + KGE
- KGE with textual enrichment (no entity linking)



1 https://en.wikipedia.org/wiki/Entity_linking

Entity Linking + KGE: Annotation



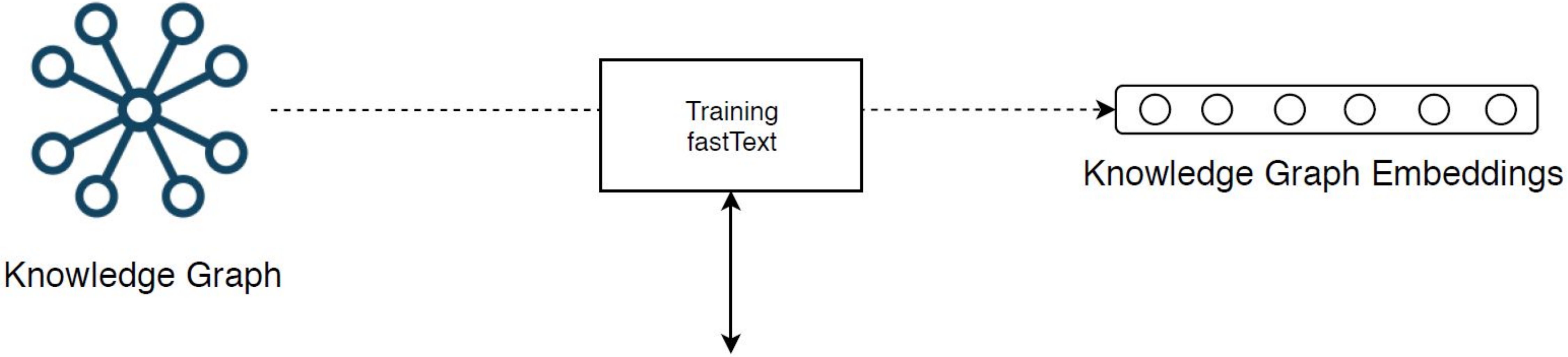
English Text

MS Paint|**dbr_Microsoft_Paint** is a good option

German Text

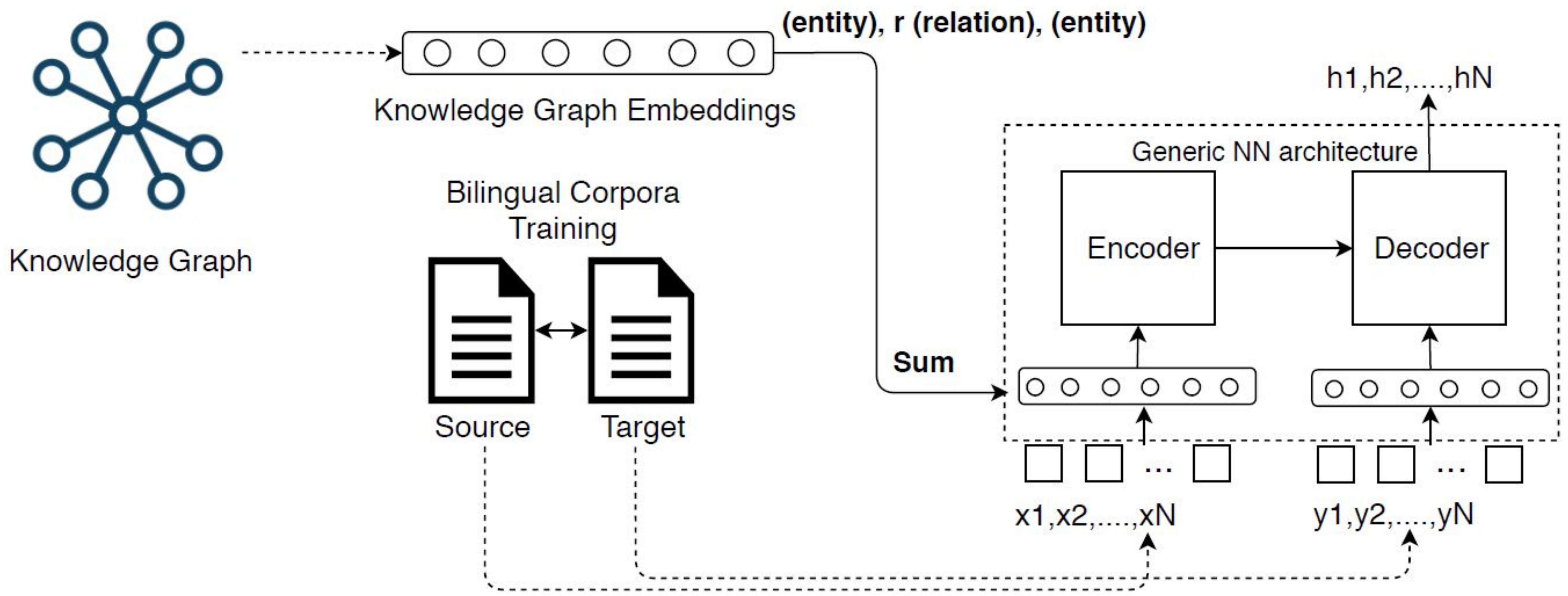
Microsoft Paint|**dbr_de_Microsoft_Paint** ist eine gute wahl

Entity Linking + KGE: KGE Training

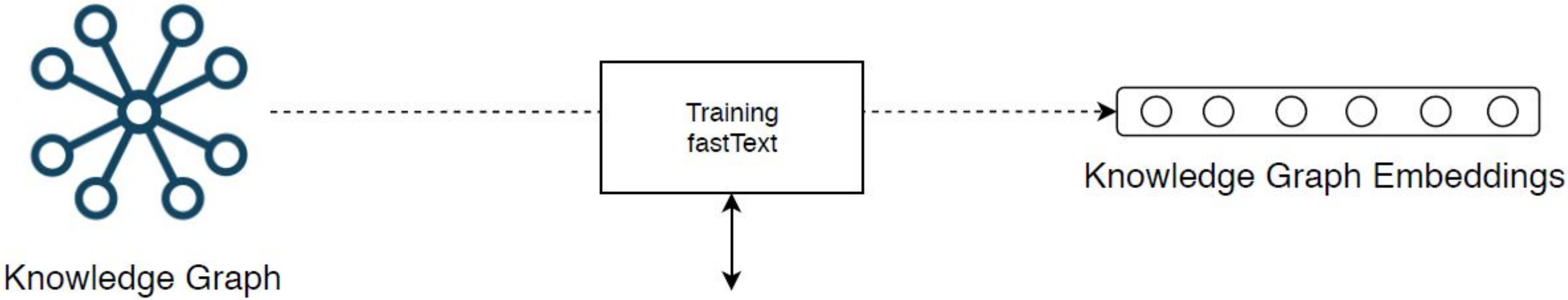


dbr_Albert_Einstein	**0_dbo_birthPlace**	dbr_Ulm
dbr_Ulm	**1_dbo_birthPlace**	dbr_Albert Einstein

KGE with textual enrichment (without EL)



KGE with textual enrichment



dbr_Albert_Einstein **0_dbo_birthPlace** dbr_Ulm
dbr_Ulm **1_dbo_birthPlace** dbr_Albert Einstein

+

label **dbr_Albert_Eintein** Albert Einsteil

Experimental Setup: NMT augmentation

- **Entity Linking**: MAG[1]
- **KG**: DBpedia[2]
    - English
        - 4.2M entities, 661 relations
    - German
        - 1M entities, 249 relations
- **KGE**: fastText[3]
        - Dimension: 500
        - Window size: 50 (for monolingual word embedding comparison)

1 https://agdistis.demos.dice-research.org/
2 https://wiki.dbpedia.org/
3 https://fasttext.cc/

NUI Galway
OÉ Gaillimh

Evaluation

- Evaluation Sets
  - newstest2015, newstest2016, newstest2017, newstest2018
  - ICD10[1], IFRS[2], IT2016[3]
- Evaluation Metrics
  - BLEU, METEOR, chrF3
- Training data
  - 2M English-German parallel sentences (JRC-Acquis, EuroParl, OpenSubtitles)
- Pre-Trained Embeddings on Wikipedia
  - English (9.2B words), German (1.3B words)

1 International Classification of disease
2 International Framework Reporting System
3 IT dataset / WMT

Results

- Q1: Can an NMT model augmented with a bilingual KG improve translation quality?
- Answer: Yes. Consistent improvement across all metrics

Models		newstest2015			newstest2016			newstest2017			newstest2018		
		BLEU	METEOR	chrF3	BLEU	METEOR	chrF3	BLEU	METEOR	chrF3	BLEU	METEOR	chrF3
Word-based models	biRNN-lstm baseline	16.77	35.20	41.11	18.55	36.62	42.54	15.10	33.75	39.52	20.53	39.02	43.92
	KG-NMT(EL+KGE)	19.86	38.25	42.92	22.38	40.40	45.18	18.04	36.94	41.55	24.87	43.49	46.88
	KG-NMT(SemKGE)	**21.49**	**40.19**	**44.72**	**24.01**	**42.47**	**46.84**	**19.66**	**38.89**	**43.11**	**27.02**	**45.77**	**48.70**
CopyM models	biRNN-lstm baseline	19.63	39.20	46.38	21.37	40.90	47.85	17.88	37.89	44.85	24.22	43.96	50.15
	KG-NMT(EL+KGE)	22.46	41.67	48.28	25.05	44.23	50.66	20.77	40.58	47.04	28.44	47.86	53.25
	KG-NMT(SemKGE)	**24.08**	**43.43**	**49.72**	**26.70**	**46.08**	**52.05**	**22.30**	**42.37**	**48.36**	**30.55**	**49.92**	**54.71**
BPE models	biRNN-lstm baseline	15.89	36.51	45.97	21.95	42.88	52.68	16.80	39.12	49.35	23.85	45.85	54.98
	KG-NMT(EL+KGE)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	KG-NMT(SemKGE)	**21.74**	**41.41**	**50.04**	**24.86**	**44.32**	**53.59**	**20.45**	**40.62**	**49.45**	**28.02**	**47.51**	**55.16**

Results

- Q2: Which method performs best on NMT, supervised annotation with EL + KGEs or unsupervised annotation with only semantically-enriched KGEs?
- Answer: unsupervised annotation with only semantically-enriched KGEs
- However:
  - Another Entity Linker can influence the results of the first strategy
  - Different KGEs can empower the second strategy
  - First strategy took five more days to train than the second one

Results

- Q3: Are KGEs more effective than pre-trained monolingual word embeddings for improving the translation of entities in NMT?
- Answer: Yes, although both presented similar scores, KG-NMT translated correctly the targeted entities

Models		newtest2015			newtest2016			newtest2017			newtest2018		
		BLEU	METEOR	chrF3	BLEU	METEOR	chrF3	BLEU	METEOR	chrF3	BLEU	METEOR	chrF3
Word-based models	biRNN-lstm+MonoE	**21.59**	**40.54**	**45.37**	**24.12**	**42.82**	**47.37**	**20.05**	**39.42**	**43.90**	**27.15**	**46.13**	**49.35**
	KG-NMT(SemKGE)	21.49	40.19	44.72	24.01	42.47	46.84	19.66	38.89	43.11	27.02	45.77	48.70
CopyM models	biRNN-lstm+MonoE	**24.21**	**43.81**	**50.32**	**26.97**	**46.52**	**52.61**	**22.61**	**42.87**	**49.01**	**30.77**	**50.39**	**55.41**
	KG-NMT(SemKGE)	24.08	43.43	49.72	26.70	46.08	52.05	22.30	42.37	48.36	30.55	49.92	54.71
BPE Models	biRNN-lstm+MonoE	19.65	39.24	47.58	**25.13**	**44.66**	**53.54**	**20.93**	**41.41**	**50.33**	**28.42**	**48.00**	**55.98**
	KG-NMT(SemKGE)	**21.74**	**41.41**	**50.04**	24.86	44.32	52.59	20.45	40.62	49.45	28.02	47.51	55.16

Results

- Q4: Can KGEs be employed for translating domain-specific knowledge
- Answer: Yes, but requires further investigation

Models		ICD-10			IFRS			IT		
		BLEU	METEOR	chrF3	BLEU	METEOR	chrF3	BLEU	METEOR	chrF3
Word-based models	biRNN-lstm baseline_adapt	15.31	23.27	29.63	**52.59**	**60.59**	**62.04**	11.57	28.04	30.50
	KG-NMT(EL+KGE)_adapt	**21.08**	**31.07**	36.93	52.38	60.55	61.86	21.78	40.29	42.75
	KG-NMT(SemKGE)_adapt	20.79	30.70	**37.00**	51.58	59.18	60.05	**23.41**	**41.71**	**44.42**
CopyM models	biRNN-lstm baseline_adapt	16.59	26.49	39.12	**52.91**	**61.68**	**64.34**	13.87	31.61	36.10
	KG-NMT(EL+KGE)_adapt	**22.59**	**34.54**	**46.89**	52.72	61.97	64.65	25.31	44.24	48.96
	KG-NMT(SemKGE)_adapt	22.24	34.10	46.74	51.91	60.33	62.51	**26.84**	**45.81**	**50.38**
BPE Models	biRNN-lstm baseline_adapt	**41.98**	**50.81**	**66.87**	**66.74**	**74.83**	**84.52**	27.83	47.01	**55.91**
	KG-NMT(EL+KGE)_adapt	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	KG-NMT(SemKGE)_adapt	41.44	50.54	66.12	66.21	74.45	84.07	**28.04**	**47.30**	55.68

Summary

- We showed:
  - incorporation of KGs into NMT models
  - ~3 points consistent improvements in BLEU, METEOR and chrF3 metric
- Next:
  - usage of other KGs such as ConceptNet, Princeton WordNet and WikiData, as well as automatically generated KGs from a set of documents
  - apply KG-NMT on low-resourced languages and scenarios

Thanks

Questions?

Mihael Arcan

Unit for Natural Language Processing

Insight SFI Research Centre for Data Analytics

Data Science Institute, NUI Galway

mihael.arcan@nuigalway.ie