Large Scale QA-SRL Parsing

Nicholas FitzGerald, Julian Michael, Luheng He, and Luke Zettlemoyer

ACL 2018



http://qasrl.org/





Subject	Manner	Verb	Object	Time
John	surreptitiously	ate	the burrito	at 2am.



• Applied to improve state-of-the-art in NLP tasks such as Question Answering [Shen 2007] and Machine Translation [Liu and Gildea, 2010]



- Applied to improve state-of-the-art in NLP tasks such as Question Answering [Shen 2007] and Machine Translation [Liu and Gildea, 2010]
- Commonly used interface to facilitate Data Exploration and Information Extraction [Stanovsky et al 2018] [Chiticariu et al. 2018]



- Applied to improve state-of-the-art in NLP tasks such as Question Answering [Shen 2007] and Machine Translation [Liu and Gildea, 2010]
- Commonly used interface to facilitate Data Exploration and Information Extraction [Stanovsky et al 2018] [Chiticariu et al. 2018]
- Considerable interest in general-purpose SRL parsers



[He et al. 2015]



[He et al. 2015]

Goal

A high-quality, large-scale parser for QA-SRL

In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

	When was something published?	In 1950
	Who published something?	Alan M. Turing
published	published "Control What was published?	
	Where was something published?	in Mind
	When did someone propose something?	In 1950
	Who proposed something?	Alan M. Turing
proposed	What did someone propose?	that machines could be tested for intelligent using questions and answers
	What did someone propose something in?	"Computing Machinery and Intelligence"
	What can be tested?	machines
tested	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
	What was being used?	questions and answers
using	Why was something being used?	tested for intelligence

1. Scale up QA-SRL data annotation

1. Scale up QA-SRL data annotation

Pierre Vinken, 61 years old, will join the	
board as a nonexecutive director Nov. 29.	
Who will join something?	
Pierre Vinken	
Question about "join" (+5c)	
What will someone join?	
When will someone join something?	
Where will someone join something?	
Why will someone join something? How How long	
How much What When Where W	of France. om the kingdom of Congo to the king of France the kingdom of Congo

1. Scale up QA-SRL data annotation



1. Scale up QA-SRL data annotation

2. Train a QA-SRL Parser



		What produced something?	A much larger super eruption
	Produced	Where did something produce something?	in Colorado
		What did something produce?	over 5,000 cubic kilometers of material
		Where didn't someone appear to do something?	In the video
		Who didn't appear to do something?	the perpetrators
la tha side a tha	appeared	When did someone appear?	never
In the video, the perpetrators never			look at the camera
appeared to look at the camera.		What didn't someone appear to do?	to look at the camera
at the camera.		Where didn't someone look at something?	In the video
	look	Who didn't look?	the perpetrators
		What didn't someone look at?	the camera
			Some of the vegetarians
		Who met someone?	vegetarians
Some of the vegetarians	met	Who met?	he
he met were members of the Theosophical		What did someone meet?	members of the Theosophical Society
Society, which had been		What had been founded?	members of the Theosophical Society
founded in 1875 to further universal		what had been founded?	the Theosophical Society
brotherhood, and which	founded	When was comothing founded?	in 1875
was devoted to the study of Buddhist and		When was something founded?	1875
Hindu literature.		Why has something been founded?	to further universal brotherhood
		Mile at the end of the data and the second state of the	we we have a fully a Theorem which had Openiated

1. Scale up QA-SRL data annotation

2. Train a QA-SRL Parser

3. Improve Recall



In 1668 the king of Portugal <u>sent</u> an elephant from the kingdom of Congo to <mark>the king of France</mark>.

+11% data

+ 2% Fscore

Invalid Where did someone send something? from the kingdom of Congo to the king of France Invalid Where did someone send something from? the kingdom of Congo Invalid Who did someone send something to?--> the king of France / Highlight to add an answer

3onus: 4c





In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

	When was something published?	In 1950
	Who published something?	Alan M. Turing
published	published "Cor What was published? a	
	Where was something published?	in Mind
	When did someone propose something?	In 1950
	Who proposed something?	Alan M. Turing
proposed	What did someone propose?	that machines could be tested for intelligent using questions and answers
	What did someone propose something in?	"Computing Machinery and Intelligence"
	What can be tested?	machines
tested	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
	What was being used?	questions and answers
using	Why was something being used?	tested for intelligence

Large-scale QA-SRL Parsing

- 1. Scale up QA-SRL data annotation
- 2. Train a QA-SRL Parser
- 3. Improve Recall

Easier Annotation

UCCA [Abend and Rapaport 2013]	~6k sentences	4 Trained Annotators
Semantic Proto-roles [Reisinger et al. 2015]	~7k sentences	MTurk
Groningen Meaning Bank [Basile et al. 2012]	~40k sentences	Trained annotators/ GWAP
QASRL 1.0 [He et al. 2015]	~3k sentences	Trained annotators

Questions:

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How						

John surreptitiously ate the burrito at 2am.

Questions:

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How						

John surreptitiously ate the burrito at 2am.

Questions:

Who ate something?

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How						

John surreptitiously ate the burrito at 2am.

Questions:

Who ate something?

What did someone eat?

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How						

John surreptitiously ate the burrito at 2am.

Questions:

. . .

Who ate something?

What did someone eat?



John surreptitiously ate the burrito at 2am.

Questions:

. . .

Who ate something?

What did someone eat?

Answers:

John the burrito

Annotation Pipeline

Predicate detection

 \mathcal{X}

John surreptitiously ate the burrito at 2am J
Identify verbs with POS + heuristics

Annotation Pipeline



Annotation Pipeline



- Efficiency
- Recall

- Efficiency
 - Autocomplete

Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.

Who will join							
Who will join as							
Who will join by							
Who will join do							
Who will join doing							
Who will join for							
Who will join from							
Who will join in							
Who will join someone							
Who will join something							
Who will join somewhere							
Who will join to							
Who will join with							

• Efficiency

- Autocomplete
- Recall
 - Autosuggest

Pierre Vinken, 61 years old, will join the board as a nonexecutive director Nov. 29.

	Who will join something?					
	Pierre Vinken					
	Question about "join" (+5c)					
	What will someone join?					
	When will someone join something?					
	Where will someone join something?					
	Why will someone join something?					
	How					
	How long					
	How much	10				
1	What					
	When					
	Where					

- Efficiency
 - Autocomplete
- Recall
 - Autosuggest
 - Financial Incentives

Validation Interface

In 1668 the king of Portugal <u>sent</u> an elephant from the kingdom of Congo to the king of France.

Invalid Where did someone send something? from the kingdom of Congo to the king of France Invalid Where did someone send something from? the kingdom of Congo Invalid Who did someone send something to?--> the king of France / Highlight to add an answer

Bonus: 4c

Dataset

- I annotator provides questions
- 2 annotators validate -> 3 spans / question
 - Question invalid if any annotator marks invalid

• Additional 3 validators for small dense dev and test set

Dataset

	Wikipedia	Wikinews	Science
Sentences	15,000	14,682	46,715
Verbs	32,758	34,026	66,653
Questions	75,867	80,081	143,388
Valid Qs	67,146	70,555	127,455

[He et al 2015]	This work	
3000 sentences	75k sentences	
Several weeks	9 days	
~50c / verb	33c / verb	
2.43 questions / verb	2.05 questions / verb	

In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

published	Who published something?	Alan M. Turing
	What was published?	"Computing Machinery and Intelligence"
	When was something published?	In 1950
	Where was something published?	in Mind
proposed	Who proposed something?	Alan M. Turing
	What did someone propose?	that machines could be tested for intelligent using questions and answers
	When did someone propose something?	In 1950
	What did someone propose something in?	"Computing Machinery and Intelligence"
tested	What can be tested?	machines
	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
using	What was being used?	questions and answers
	Why was something being used?	tested for intelligence

In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

published	Who published something?	Alan M. Turing
	What was published?	"Computing Machinery and Intelligence"
	When was something published?	In 1950
	Where was something published?	in Mind
proposed	Who proposed something?	Alan M. Turing
	What did someone propose?	that machines could be tested for intelligent using questions and answers
	When did someone propose something?	In 1950
	What did someone propose something in?	"Computing Machinery and Intelligence"
tested	What can be tested?	machines
	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
using	What was being used?	questions and answers
	Why was something being used?	tested for intelligence
In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

	Who published something?	Alan M. Turing	
published	What was published?	"Computing Machinery and Intelligence"	
	When was something published?	In 1950	
	Where was something published?	in Mind	
proposed	Who proposed something?	Alan M. Turing	
	What did someone propose?	that machines could be tested for intelligent using questions and answers	
	When did someone propose something?	In 1950	
	What did someone propose something in?	"Computing Machinery and Intelligence"	
tested	What can be tested?	machines	
	What can something be tested for?	intelligence	
	How can something be tested?	using questions and answers	
using	What was being used?	questions and answers	
	Why was something being used?	tested for intelligence	

Large-scale QA-SRL Parsing

- 1. Scale up QA-SRL data annotation
- 2. Train a QA-SRL Parser
- 3. Improve Recall

QA-SRL Parser



QA-SRL Parser



QA-SRL Parser

- Unlabeled Argument Detection
- Question generation

 ${\mathcal X}$ John surreptitiously ate the burrito at 2pm



Predicate detection











Argument Detection - BIO Model

- Alternating Bi-LSTM with Highway Connections and Recurrent Dropout [He et al 2017]
- Input includes predicate indicator



Argument Detection - BIO Model



Argument Detection - BIO Model















Argument Detection

- 4 layer Alternating Bi-LSTM with Highway Connections and Recurrent Dropout [He et al 2017]
- Trained to maximize log-likelihood

Argument Detection



Question Generation

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How					• • •	

Question Generation

Wh	Aux	Subj	Verb	Obj	Prep	Obj2
Who	Ø	Ø	Ø	Ø	Ø	Ø
What	did	someone	stem	someone	on	someone
Where	didn't	something	past	something	to	something
When	might		past participle		by	
Why	will		present		from	
How					•••	
				- 1 1 1 1 1 1 1		

- Two models:
 - Local slot prediction
 - Sequential Model (LSTM)









Question Generation - Sequential



Question Generation - Sequential



Question Generation - Sequential



Question Generation

- 4 layer Alternating Bi-LSTM with Highway Connections and Recurrent Dropout [He et al 2017]
- Trained to maximize log-likelihood

Evaluation Questions

• Paraphrasing

"Who ate something?

"Who was something eaten by?

Evaluation Questions

• Paraphrasing

"Who ate something?

"Who was something eaten by?

- Given gold span:
 - Slot-level accuracy
 - Exact Match (full question)

Question Generation


Full Parsing Accuracy



Exact match f-score (Span & Question)

Large-scale QA-SRL Parsing

- 1. Scale up QA-SRL data annotation
- 2. Train a QA-SRL Parser
- 3. Improve Recall

In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

published	Who published something?	Alan M. Turing	
	What was published?	"Computing Machinery and Intelligence"	
	When was something published?	In 1950	
	Where was something published?	in Mind	
proposed	Who proposed something?	Alan M. Turing	
	What did someone propose?	that machines could be tested for intelligent using questions and answers	
	When did someone propose something?	In 1950	
	What did someone propose something in?	"Computing Machinery and Intelligence"	
tested	What can be tested?	machines	
	What can something be tested for?	intelligence	
	How can something be tested?	using questions and answers	
using	What was being used?	questions and answers	
	Why was something being used?	tested for intelligence	

In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

	Who published something?	Alan M. Turing
published	What was published?	"Computing Machinery and Intelligence"
	When was something published?	In 1950
	Where was something published?	in Mind
	Who proposed something?	Alan M. Turing
proposed	npacts training an	that machines could be tested for intelligent using questions and answers
		In 1950
	When did someone propose something? Contract of the someone propose something in?	"Computing Machinery and Intelligence"
	What can be tested?	machines
tested	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
	What was being used?	questions and answers
using	Why was something being used?	tested for intelligence

In 1950 Alan M. Turing *published* "Computing machinery and intelligence" in Mind, in which he *proposed* that machines could be *tested* for intelligence *using* questions and answers.

	Who published something?	Alan M. Turing
published	What was published?	"Computing Machinery and Intelligence"
	When was something published?	In 1950
	Where was something published?	in Mind
	Who proposed something?	Alan M. Turing
proposed	What did someone propose?	that machines could be tested for intelligent using questions and answers
	Fillin Gaps	In 1950
	What did someone propose something in?	"Computing Machinery and Intelligence"
	What can be tested?	machines
tested	What can something be tested for?	intelligence
	How can something be tested?	using questions and answers
	What was being used?	questions and answers
using	Why was something being used?	tested for intelligence

Annotation Pipeline



Annotation Pipeline



Data Expansion

- Overgenerate questions with low span threshold
- Validate
- +46,715 valid QA-pairs (~11% after filtering)

Data Expansion



Evaluation

- Exact Match for Question Generation is overly harsh
 - Paraphrasing

"Who ate something?

"Who was something eaten by?

- Missing Questions
 - Penalizes correct predictions missing from data

Human Evaluation

- Validate model predictions with 6 annotators
- Generated Question valid if 5 out of 6 annotators provided answers
- Predicted span correct if exactly matches any annotator's answer

Human Eval - Questions



Human Eval - Questions



Human Eval - Arguments



Example Output

Some of the vegetarians he **met** were members of the Theosophical Society, which had been **founded** in 1875 to further universal brotherhood, and which was **devoted** to the study of Buddhist and Hindu literature.

met	Who met someone?	Some of the vegetarians	
	Who met?	he	
	What did someone meet?	members of the Theosophical Society	
founded What had been founded? the The founded When was something founded? the The	What had been founded?	members of the Theosophical Society	
	what had been founded?	the Theosophical Society	
	When was something founded?	in 1875	
	to further universal brotherhood		
devoted	What was devoted to something?	members of the Theosophical Society	
	What was something devoted to?	the study of Buddhist and Hindu literature	

Conclusion

- Large crowdsourced dataset of QA-SRL annotations
- High quality QA-SRL Parser
- Techniques for data expansion

Code and Data available at: http://qasrl.org

QA-SRL I	Bank 2.0	 Train Wikipedia Dev Wikinews 	OriginalExpansion	Legend (help): Original Expansion Eval Valid only (≤ $4/6 \rightarrow$ invalid, ≥ $5/6 \rightarrow$ valid)
Search sentences	Submit Clear	🗹 TQA	🔽 Eval	Answer provided by 1/2/3/4/5/6 annotators
783 / 5712 documents	5 sentences	Dev / TQA / T_1337 (2011 Virginia Earth	quake) / paragraph 0,	sentence 0
2011 Virginia Earthquake	In August 2011 the eastern seaboard of the U.S. was	In August 2011 the ea	astern seabo	pard of the U.S. was rocked by a magnitude 5.8 earthquake.
A Good Prediction	rocked by a magnitude 5.8 earthquake.	rocked		
Active Plate	While not huge, most of the residents had never experienced a quake and many didnt know what it	What was rocked?	3/3 the east	ern seaboard of the U.S.
Margins Active Transport		What was something rocked by?	3/3 a magni	tude 5.8 earthquake
Additional Physical Properties	was.	When was something rocked by something?	3/3 In Augus	st 2011
Advanced Topic Defining Density	Some people thought the shaking might have been the result of a terrorist attack.	How was something rocked?	3/3 by a ma	gnitude 5.8 earthquake
Advantage of Asexual Reproduction	But if you went back in time to the late Paleozoic, you would find the region being			
Advantage of Sexual Reproduction	uplifted into the ancestral Appalachian mountains as continent-continent convergence brought			
After the Big Bang	Pangaea together.			

Code and Data available at: http://demo.qasrl.org

QA-SRL Parser Demo

At ACL, Nicholas met other scientists and en	njoyed discussing research with them.	Submit	
At ACL, Nicholas met other scientists and enjoyed discussing research with them.			
met			
Where did someone meet someone?	At ACL		
Who met someone?	Nicholas		
Who did someone meet?	other scientists		
enjoyed			
Where did someone enjoy something?	At ACL		
Who enjoyed something?	Nicholas		
What did someone enjoy?	discussing research with them		
discussing			
Where was someone discussing something?	At ACL		
Who was discussing something?	Nicholas		
What was someone discussing?	research		
Who was someone discussing something with?	them		

Thank You!



http://qasrl.org/



