

# Czech Question Answering with Extended SQADv3.0 Benchmark Dataset

Marek Medveď, Aleš Horák, Radoslav Sabol

Centrum ZPJ, FI MU, Brno

December 7, 2019

addQA

# addQA system

Remaining annotations: 198 of 200

Question type	Missing QA	
<a href="#">ABBREVIATION</a>	10 of 10	Add new ABBREVIATION QA
<a href="#">ADJ_PHRASE</a>	50 of 50	Add new ADJ_PHRASE QA
<a href="#">CLAUSE</a>	10 of 10	Add new CLAUSE QA
<a href="#">DATETIME</a>	9 of 10	Add new DATETIME QA
<a href="#">ENTITY</a>	28 of 28	Add new ENTITY QA
<a href="#">LOCATION</a>	9 of 10	Add new LOCATION QA
<a href="#">NUMERIC</a>	4 of 4	Add new NUMERIC QA
<a href="#">PERSON</a>	28 of 28	Add new PERSON QA
<a href="#">VERB_PHRASE</a>	50 of 50	Add new VERB_PHRASE QA

Already annotated: 2 of 200

Question ID	Q type	A type	Question	Answer	Sentence	URL
19	LOCATION	LOCATION	Kde se nachází Kuba?	v severním Karibiku	Kuba se nachází v severním Karibiku a její břehy omývají Karibské moře, Mexický záliv a Atlantský oceán.	<a href="https://cs.wikipedia.org/wiki/Kuba">https://cs.wikipedia.org/wiki/Kuba</a>
21	DATETIME	DATETIME	Kdy se narodil Jeremy Clarkson?	11. dubna 1960	Jeremy Clarkson (* 11. dubna 1960), celým jménem Jeremy Charles Robert Clarkson, je anglický hlasatel, žurnalista a spisovatel, který se specializuje na motorismus.	<a href="https://cs.wikipedia.org/wiki/Jeremy_Clarkson">https://cs.wikipedia.org/wiki/Jeremy_Clarkson</a>

## addQA system

Question:  e.g. **Co je letadlo?**

Exact answer:  e.g. **letající dopravní prostředek**

Answer sentence(s):  e.g. **Letadlo je létající dopravní prostředek.**

Wikipedia URL:  e.g. **<https://cs.wikipedia.org/wiki/Letadlo>**

Question type:

Answer type:

## addQA system

Question:  e.g. [Co je letadlo?](#)

Retag Question:

Question tagged:  See [Tagset](#) for details.

Exact answer:  e.g. [letající dopravní prostředek](#)

Retag Answer:

Answer tagged:  See [Tagset](#) for details.

Answer sentence(s):  e.g. [Letadlo je létající dopravní prostředek.](#)

Retag Sentence(s):

Sentence(s) tagged:  See [Tagset](#) for details.

Wikipedia URL:  e.g. <https://cs.wikipedia.org/wiki/Letadlo>

Retag Text:

Full article:  See [Tagset](#) for details. The full article does not need to be checked in detail.

Question type:  See [Help](#) for details.

Answer type:  See [Help](#) for details.

# addQA system

Remaining annotations: 198 of 200

Question type	Missing QA	
<a href="#">ABBREVIATION</a>	10 of 10	Add new ABBREVIATION QA
<a href="#">ADJ_PHRASE</a>	50 of 50	Add new ADJ_PHRASE QA
<a href="#">CLAUSE</a>	10 of 10	Add new CLAUSE QA
<a href="#">DATETIME</a>	9 of 10	Add new DATETIME QA
<a href="#">ENTITY</a>	28 of 28	Add new ENTITY QA
<a href="#">LOCATION</a>	9 of 10	Add new LOCATION QA
<a href="#">NUMERIC</a>	4 of 4	Add new NUMERIC QA
<a href="#">PERSON</a>	28 of 28	Add new PERSON QA
<a href="#">VERB_PHRASE</a>	50 of 50	Add new VERB_PHRASE QA

Already annotated: 2 of 200

Question ID	Q type	A type	Question	Answer	Sentence	URL
19	LOCATION	LOCATION	Kde se nachází Kuba?	v severním Karibiku	Kuba se nachází v severním Karibiku a její břehy omývají Karibské moře, Mexický záliv a Atlantský oceán.	<a href="https://cs.wikipedia.org/wiki/Kuba">https://cs.wikipedia.org/wiki/Kuba</a>
21	DATETIME	DATETIME	Kdy se narodil Jeremy Clarkson?	11. dubna 1960	Jeremy Clarkson (* 11. dubna 1960), celým jménem Jeremy Charles Robert Clarkson, je anglický hlasatel, žurnalista a spisovatel, který se specializuje na motorismus.	<a href="https://cs.wikipedia.org/wiki/Jeremy_Clarkson">https://cs.wikipedia.org/wiki/Jeremy_Clarkson</a>

# SQAD record

- the *question*
- the *full article* text
- the *answer selection* sentence which contains the expected answer as a sub-phrase
- the *answer extraction* result, i.e. the exact expected answer
- the *URL* of the original article in the Czech Wikipedia
- the *QA metadata* – the question and answer types

## addQA system

Question type	Description	Example question	Example Answer
ABBREVIATION	The question asks for abbreviation of some name.	Jakou chemickou značku má vápník?	Ca
ADJ_PHRASE	Question asks about some specific group of things, that is usually specified by adjective.	Jaká je tradiční barva Oxfordské univerzity?	tmavě modrá
CLAUSE	Question is general and the answer can be any general clause.	Proč se chtěla Marie Terezie spojit s Francií?	Protože by Prusko ztratilo svého důležitého spojence
DATETIME	Main goal of question is to determine certain point in time.	Kdy se narodila Petr Kvitová?	8. března 1990
ENTITY	Main goal of question is to name a thing (not a person) that meets all conditions of question.	Jak se jmenuje největší planeta sluneční soustavy?	Jupiter
LOCATION	Main goal of question is to determine certain place.	Kde zemřel Josef Kajetán Tyl?	Plzeň
NUMERIC	Main goal of question is to determine certain number.	Do kolika větví je rozdělena Armáda České republiky?	Do tří
PERSON	Main goal of question is to name a person that meets all conditions of question.	Kdo byl 33. prezidentem Spojených států amerických?	Harry S. Truman
VERB_PHRASE	Main goal of question is to find out if something happened. The answer can be general verb phrase or confirmation in form of YES/NO.	Je Brno sídlem fotbalového týmu Bohemians 1905?	ne

Answer type	Description	Example question	Example Answer
ABBREVIATION	The answer is abbreviation of some name.	Jakou chemickou značku má vápník?	Ca
DATETIME	Answer is a certain point in time.	Kdy se narodila Petr Kvitová?	8. března 1990
ENTITY	Answer is a name of some thing (not a person).	Jak se jmenuje největší planeta sluneční soustavy?	Jupiter
LOCATION	Answer denotes a certain place.	Kde zemřel Josef Kajetán Tyl?	Plzeň
NUMERIC	Answer is a number.	Do kolika větví je rozdělena Armáda České republiky?	Do tří
ORGANIZATION	Answer is name of some organisation, band, company ...	Frontman jaké kapely je Jarda Svoboda?	Traband
OTHER	Answer is a general phrase that is not belong to other answer type	Co je hard rock?	hudební styl
PERSON	Answer is a name a person that meets all conditions of question.	Kdo byl 33. prezidentem Spojených států amerických?	Harry S. Truman
DENOTATION	Answer is a general name of a field, area or an approach.	Jak se nazývá obor, který se zabývá studiem chování živočichů?	Etologie
YES/NO	The answer is YES or NO.	Je Brno sídlem fotbalového týmu Bohemians 1905?	ne



# Format

# Format update

- **Goal:** automatic consistency check of SQAD database

# SQAD format update - anaphora problem

## Example (Anaphora)

*Article:*

*"... **Peter** was a famous singer. **He** was also a famous English song writer. ..."*

*Question:*

*"What is the name of a famous English song writer?"*

*Results:*

- 1 containing exact answer: *"Peter was a famous singer"*
- 2 sentence which corresponds directly to the question but the answer is just referred by the anaphora: *"He was also a famous English song writer"*
- 3 1. and 2. together

# Answer selection in SQAD v3

## Example (Anaphora)

*“Peter was a famous singer. He was also a famous English song writer.”*

- 1 The answer selection must contain exactly **one sentence**.
- 2 The answer selection sentence **must** contain the **expected answer** (as a sub-phrase).
- 3 If multiple sentences (with anaphora) are needed to uniquely identify the answer to the question, the answer sentence and the respective antecedent sentence(s) are stored in a new **“question\_context”** file.

## Answer extraction in SQAD v3

- **SQAD 2.\*** - expected answer sometimes just “followed” from the answer sentence, but it was not an exact sub-phrase of answer selection sentence
- **SQAD v3** - new “**answer\_extraction**” file now contains the exact sub-phrase of answer selection

# SQAD evaluation

# SQAD v3 statistics

	SQAD v3.0	SQAD v2.1
No. of records	<b>13,473</b>	8,566
No. of different articles	6,571	3,930
No. of tokens (words)	28,825,824	20,288,297
No. of answer contexts	378	0

# Q-type statistics

Q-Type statistics:	SQAD v3.0	SQAD v2.1
DATETIME	14.7 %	21.6 %
PERSON	13.1 %	11.9 %
VERB_PHRASE	<b>16.8 %</b>	10.97 %
ADJ_PHRASE	<b>11.2 %</b>	2.7 %
ENTITY	18.4 %	20.4 %
CLAUSE	3.5 %	2.8 %
NUMERIC	7.3 %	10.7 %
LOCATION	12.4 %	17.8 %
ABBREVIATION	<b>2.5 %</b>	0.95 %
OTHER	0.1 %	0.18 %

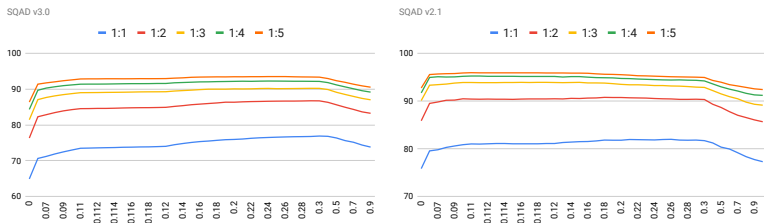


# Q-type statistics

A-Type statistics:	SQAD v3.0	SQAD v2.1
DATETIME	14.6 %	21.5 %
PERSON	13.2 %	12.3 %
YES_NO	16.8 %	10.95 %
OTHER	16.7 %	9.6 %
ENTITY	13.1 %	12.7 %
NUMERIC	7.4 %	10.7 %
LOCATION	12.3 %	17.6 %
ABBREVIATION	2.4 %	0.96 %
ORGANIZATION	2.1 %	2.5 %
DENOTATION	1.4 %	1.2 %

# AQA evaluation

# Evaluation: AQA document selection level



**Figure:** Comparison of SQAD v2.1 and SQAD v3.0 on the document selection level (combined score on top 5 documents)


# Evaluation: AQA answer selection level

Training answer selection model:

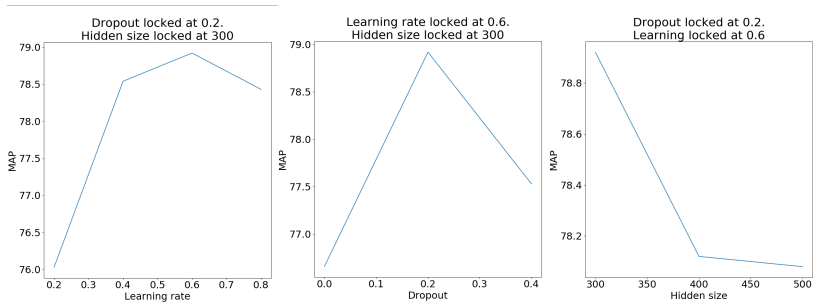
- SQuAD v3 (balanced):  $60_{train} : 10_{val} : 30_{test}$
- hidden size: **300**, 400, 500
- learning rate: 0.2, 0.4, **0.6**, 0.8
- dropout: 0, **0.2**, 0.4
- 25 epochs
- optimizer: Stochastic Gradient Descent
- together 36 combinations, evaluated 3 times → **108** models
- Mean Average Precision<sup>1</sup>: **78.92%**
- Mean Reciprocal Rank<sup>2</sup>: **85.95**

---

<sup>1</sup>how good model is at performing the query

<sup>2</sup>average of the reciprocal ranks of results for a sample of queries Q 

# Evaluation: AQA answer selection level



**Figure:** The sensitivity graphs of the hidden size, dropout and learning rate hyperparameters in the SQADv3 evaluation.

# Evaluation: AQA answer selection level

Table: The SQuADv3 answer selection accuracy per question types

Question type	Count	MAP (%)
VERB_PHRASE	546	80.06
NUMERIC	212	74.13
ADJ_PHRASE	363	79.08
CLASUE	99	67.81
DATETIME	473	82.12
ABBREVIATION	71	78.89
LOCATION	417	<b>84.76</b>
ENTITY	571	77.16
PERSON	418	77.41
OTHER	1	<b>33.33</b>

# Evaluation: AQA answer selection level

Table: The SQADv3 answer selection accuracy per answer types

Answer type	Count	MAP (%)
YES_NO	539	79.73
NUMERIC	215	<b>73.63</b>
OTHER	526	76.79
DATETIME	470	81.88
LOCATION	415	84.87
ENTITY	403	76.47
PERSON	421	77.53
ABBREVIATION	66	78.57
ORGANIZATION	65	75.58
DENOTATION	51	<b>87.93</b>

# Evaluation: AQA answer selection level

Table: Answer selection Precision at  $k$  ( $P@k$ )

Position $k$	Count	$P@k$	Position $k$	Count	$P@k$
1.	3,171	79.00%	6.	31	0.77%
2.	377	9.39%	7.	20	0.50%
3.	125	3.11%	8.	16	0.40%
4.	66	1.64%	9.	13	0.32%
5.	59	1.47%	$\geq 10.$	136	3.40%
		94.61			



# Evaluation: AQA answer selection level

**Table:** Results of 5-fold cross validation for both versions of the SQAD dataset

Test partition no.	SQADv3.0 MAP	SQADv2.1 MAP
1	82.76%	87.16%
2	74.26%	81.28%
3	71.98%	76.20%
4	80.85%	73.85%
5	78.55%	67.35%
overall	<b>77.68%</b>	<b>77.17%</b>

## Conclusion and future work

- SQAD v3 (13,473 records)
- addQA (add all restrictions)
- AQA evaluation

# Acknowledgement

# Acknowledgement

This work has been partly supported by the Czech Science Foundation under the project GA18-23891S.

Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme “Projects of Large Research, Development, and Innovations Infrastructures” (CESNET LM2015042) is greatly appreciated.