

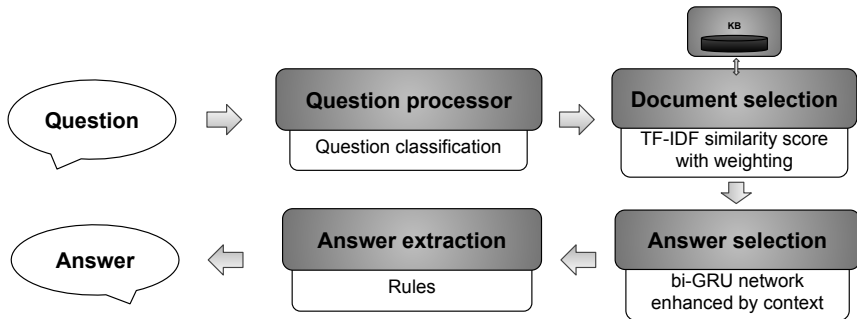
# Evaluating Long Contexts in the Czech Answer Selection Task

Marek Medved' Radoslav Sabol Aleš Horák  
xmedved1@fi.muni.cz, xsabol@fi.muni.cz, haless@fi.muni.cz

NLP Centre, Faculty of Informatics, Masaryk University

December 10, 2021

# AQA structure



## Answer context - Motivation

---

### Question:

*Kolik členů má v současnosti finská hudební skupina **Apocalyptica**?  
[How many members does the Finnish band Apocalyptica currently have?]*

### Correct answer:

*Skupina je složena ze tří (původně čtyř) klasických violoncellistů.  
[The band is composed of three (originally four) classical cellists.]*

### Answer context (previous sentence):

***Apocalyptica** je finská hudební skupina, jejíž zvláštností je interpretace původně heavy metalových skladeb osobitým způsobem aranžovaných provioloncello.*

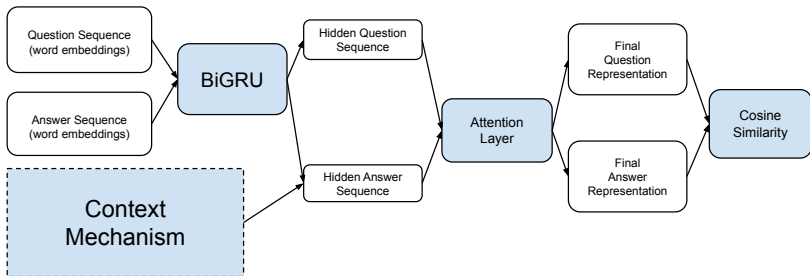
*[Apocalyptica is a Finnish band whose peculiarity is the interpretation of originally heavy metal compositions arranged in a special way for cello.]*

---

## Answer Context - Intuition

- question contains important clues
- clue information is often distributed not only in the final answer sentence but also in the preceding text
- The longer the preceding answer context is, the more precise and certain the sought answer is. - common assumption of people
- can features/knowledge obtained from answer context help the final model to identify the correct answer more precisely?

# Answer selection structure

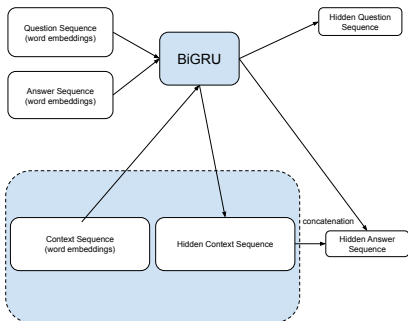


## Context kinds

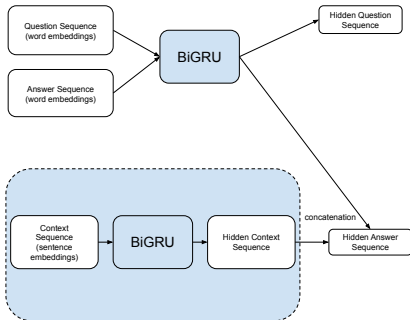
- **RNN based:** offers a direct extension of the candidate answer sentence serving the context as input to the same recurrent network (RNN) layer as used for the answer
- **transformer based:** uses various pre-trained transformer representations of the whole preceding sentence (or sentences) as the context

# Answer selection context

## RNN Context



## Transformer Context



## Context types

- previous sentence
- noun phrase
- linked named entities



## Context length

Context type	context window (sentences)	average length of context in tokens	average number of context items
NER	1	2.29	1.49
	2	4.48	2.97
	3	6.70	4.45
	4	8.93	5.93
	5	11.16	7.41
PHR	1	13.77	5.08
	2	27.71	10.22
	3	41.55	15.33
	4	55.42	20.45
	5	69.30	25.58
SENT	1	19.97	1.00
	2	40.12	2.00
	3	60.24	3.00
	4	80.41	4.00
	5	100.60	5.00

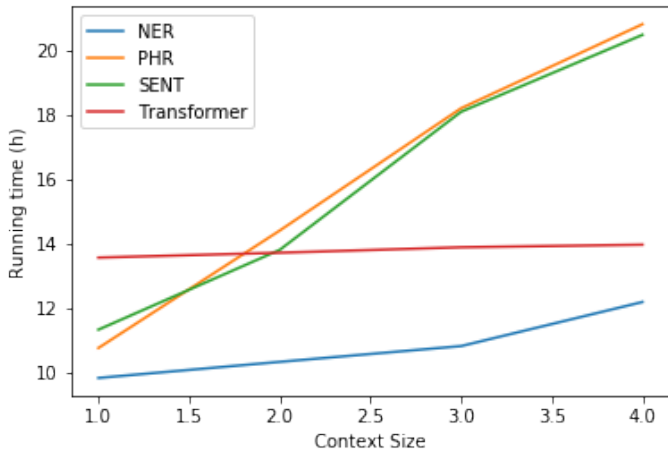
# Models

- FastText
- Bert based
  - Czert
  - Slavic Bert
  - RobeCzech

# Experiment Setup

- *Which context type works the best?*
  - evaluation of PHR, NER, SENT (both RNN and Transformer) contexts
- *What kind of influence does context window size have on the results?*
  - context window - a number of preceding sentences to extract the context from
  - 1-4 used for the following experiments

## Running times of experiments with respect to the context type and window



## Mean average precision for each context type and context window size

CT	Mean Average Precision							
	1		2		3		4	
W	S	M	S	M	S	M	S	M
MAP								
PHR	82.24	84.92	82.23	84.98	80.56	83.41	80.55	83.31
NER	82.58	85.3	82.16	84.94	82.71	85.53	82.4	85.04
SENT	81.9	84.76	80.9	83.39	79.31	82.2	78.54	81.56
CZERT	<b>83.39</b>	<b>85.79</b>	82.71	85.38	82.76	85.36	82.78	85.35
RBCZ	82.75	85.29	82.46	85.05	82.69	85.44	82.56	85.14
SLAV_B	83.05	85.59	83.19	85.91	82.74	85.49	82.88	85.55

## Best models per question type with different context types

Question type	Non context MAP in %	best context	window	best MAP in %
VERB_PHRASE	82.64	NE	3	<b>83.63</b>
ENTITY	79.40	SLB	1	<b>81.62</b>
NUMERIC	78.50	NE	1	<b>79.79</b>
ADJ_PHRASE.	83.89	SLB	1	<b>84.19</b>
CLAUSE	74.82	SLB	2	<b>75.78</b>
DATETIME	84.52	CZT	1	<b>84.80</b>
LOCATION	83.13	CZT	1	<b>86.61</b>
PERSON	81.33	CZT	1	<b>85.17</b>
ABBREVIATION	91.75	NE	4	<b>94.16</b>

## NER example

---

**Question:**

*Kolik sportovců se zúčastnilo XXVIII. letních olympijských her 2004 textitv Aténách?*

*[How many athletes participated in the XXVIII-th Summer Olympic games in 2004 in Athens?]*

**Answer from non-context model:**

*Her se zúčastnilo 202 zemí.*

*[202 countries took part in the games.]*

**Answer from the NER context model (window size of one sentence):**

*Účastnilo se jich 10625 sportovců z 201 zemí světa.*

*[10625 athletes from 201 countries took part in them. ]*

**1th context item**

*letní olympijské hry*

*[Summer Olympic games]*

**2nd context item**

*Athénách*

*[Athens]*

---

## SENT example

---

**Question:**

*Je Jeruzalém jedno z nejstarších měst na světě?*

*[Is Jerusalem one of the oldest cities in the world?]*

**Answer from the non-context model:**

*Historie města sahá až do 4. tisíciletí př. n. l. a činí tak z Jeruzaléma jedno z nejstarších měst na světě.*

*[The history of the city dates back to the 4-th millennium BC and makes Jerusalem one of the oldest cities in the world.]*

---



## SENT example cont.

---

**Answer from the SENT context model (window size of 4 previous sentences):**

*Nachází se v něm však také množství významných starověkých křesťanských míst a je považováno za třetí nejsvětější místo islámu. [However, there us also located a number of important ancient Christian sites and is considered the third holiest site in Islam.]*

**1th context item:**

*Jeruzalém se nachází v Judských horách na hranici úmoří Středozevního a Mrtvého moře na okraji Judské pouště.*

**2nd context item:**

*Současný Jeruzalém se rozrůstá daleko za hranicemi Starého Města.*

**3rd context item:**

*Historie města sahá až do 4. tisíciletí př. n. l. a činí tak z Jeruzaléma jedno z nejstarších měst na světě.*

**4th context item:**

*Jeruzalém je nejsvětějším místem judaismu a duchovním centrem židovského národa.*

---

## Conclusion

- transformer-based sentence context is the most compact and efficient representation
  - generally, more compact representations yielded better results
- increasing the context size beyond 1 rarely leads to an improvement

**MUNI**

FACULTY

OF INFORMATICS