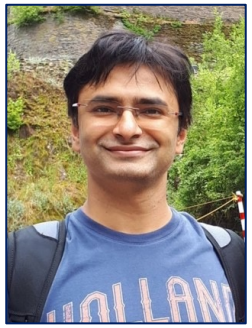
Beyond NED:
Fast and Effective Search Space Reduction for Complex Question Answering over Knowledge Bases

Philipp Christmann, *Rishiraj Saha Roy*, *Gerhard Weikum*



SIC Saarland Informatics Campus

max planck institut informatik

Standard KB-QA pipeline

Who scored in the 2018 final between France and Croatia?

Curated **knowledge bases** (KB) have
⇒ Billions of facts
⇒ Millions of entities
⇒ Thousands of relations

Question Answering (QA) system typically neural model
⇒ Neural reader, graph neural network,…
⇒ Requires sufficiently small input

⇒ QA system **not directly applicable** to **full KB**

Standard KB-QA pipeline

Who scored in the 2018 final between France and Croatia?

Mario Mandžukić, Ivan Perišić, Antoine Griezmann, Paul Pogba

Question Answering System

Search space S
(Set of KB-facts)

Search Space Reduction

Standard KB-QA pipeline

Who scored in the 2018 final between France and Croatia?

Named-entity recognition and disambiguation (NERD)

"2018 final" ↦ 2018 EuroLeague Final
"France" ↦ France football team
"Croatia" ↦ Croatia football team

Fact Retrieval
from Knowledge Base

Search space S
(Set of KB-facts)

Question Answering System

?

Beyond NERD

Who **scored** in the 2018 final between France and Croatia?

Top-1 NERD too restrictive

⇒ **Top-*k*** for all entities can include **noise** and **blows up search space**

⇒ **Adaptive Top-*k*:** *k* **individually** per question word

NERD not tailored for QA

⇒ Disambiguate **all KB-items**

- ⇒ **Entities**
- ⇒ **Types** (e.g. National football team, FIFA World Cup Final)
- ⇒ **Concepts** (e.g. football, goal)
- ⇒ **Predicates** (e.g. goal scored by, location)

Disambiguations ideally establish a **compact KB-subgraph**

⇒ Consider **proximity** of disambiguations in graph underlying KB

CLOCQ

KB-QA with CLOCQ

Who scored in the 2018 final between France and Croatia?

Mario Mandžukić, Ivan Perišić, Antoine Griezmann, Paul Pogba

Question Answering System

"scored"	↦	goal scored by no. of goals scored
"2018 final"	↦	2018 FIFA WC Final 2018 EuroLeague Final 2018 ChampionsLeague Final 2018 FA Cup Final
"France"	↦	France football team France basketball team France (state)
"Croatia"	↦	Croatia football team Croatia basketball team Croatia (state)

Search space S
(Set of KB-facts)

CLOCQ

KB-item retrieval

Who scored in the 2018 final between France and Croatia?

⇒ Retrieve ***d* candidate KB-items** for each question word (same *d* for all)

⇒ **Gold KB-items** can be **deep down** in the list

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
...	...
5	2018 FA Cup Final
6	2018 NBA Finals
...	...

France

	KB item
1	France (state)
2	Kingdom of France
...	...
21	France basketball team
22	**France football team**
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
25	**Croatia football team**
26	Croatia basketball team
...	...

Adaptive Top-*k*

Who scored in the 2018 final between France and Croatia?

⇒ Retrieve ***d* candidate KB-items** for each question word (same *d* for all)

⇒ **Gold KB-items** can be **deep down** in the list

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
...	...

***k* = 2**

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
...	...
5	2018 FA Cup Final
6	2018 NBA Finals
...	...

***k* = 4**

France

	KB item
1	France (state)
2	Kingdom of France
...	...
21	France basketball team
22	**France football team**
...	...

***k* = 3**

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
25	**Croatia football team**
26	Croatia basketball team
...	...

***k* = 3**

Auto-*k* mechanism

⇒ Choose **different *k* for each word**

⇒ Mechanism based on **ambiguity**

Scoring KB-items

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

Scoring KB-items

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...

X

⇒ **Cross product** of KB items
⇒ Computationally **not feasible**

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

Candidate scoring

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	goal scored by
4	film score
...	...

	KB item	score
1	goal scored by	0.58
2	no. of goals scored	0.52

$k = 2$

Rank

Aggregation

l_{12}	KB item	conn
1	goal scored by	0.66
2	no. of goals scored	0.50
3	film score	0.43
...	...	...

Pair-wise KB-proximity

l_{11}	KB item	coh
1	no. of goals scored	0.93
2	goal scored by	0.91
3	score (music)	0.81
...	...	...

Pair-wise semantic coherence

l_{13}	KB item	rel
1	goal scored by	0.76
2	score (music)	0.71
3	no. of goals scored	0.65
...	...	...

Question relevance

l_{14}	KB item	match
1	score (music)	0.50
2	no. of goals scored	0.33
3	goal scored by	0.25
...	...	...

Lexical matching

score = h_{coh} coh + h_{conn} conn + h_{rel} rel + h_{match} match

KB-promixity

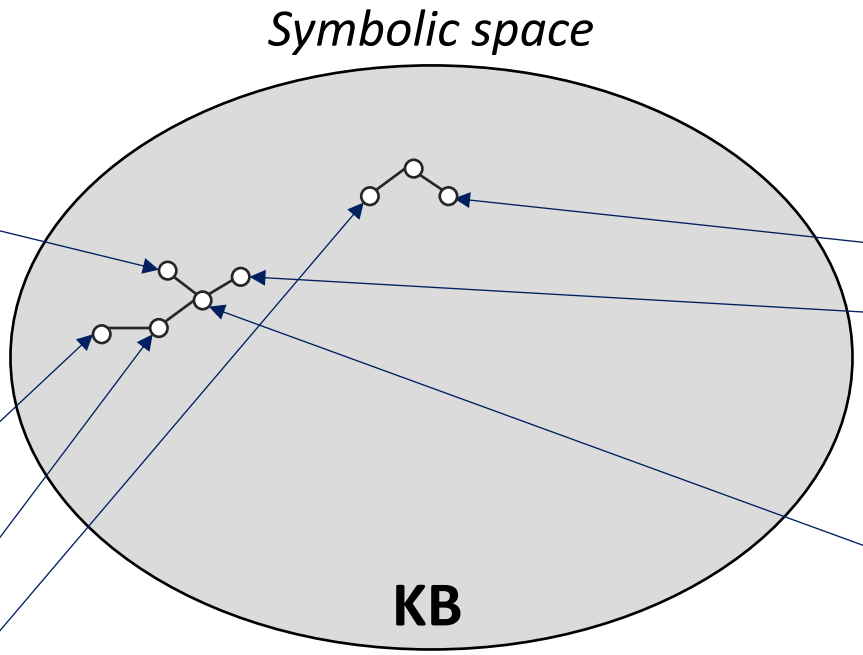
Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...



France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

KB-promixity

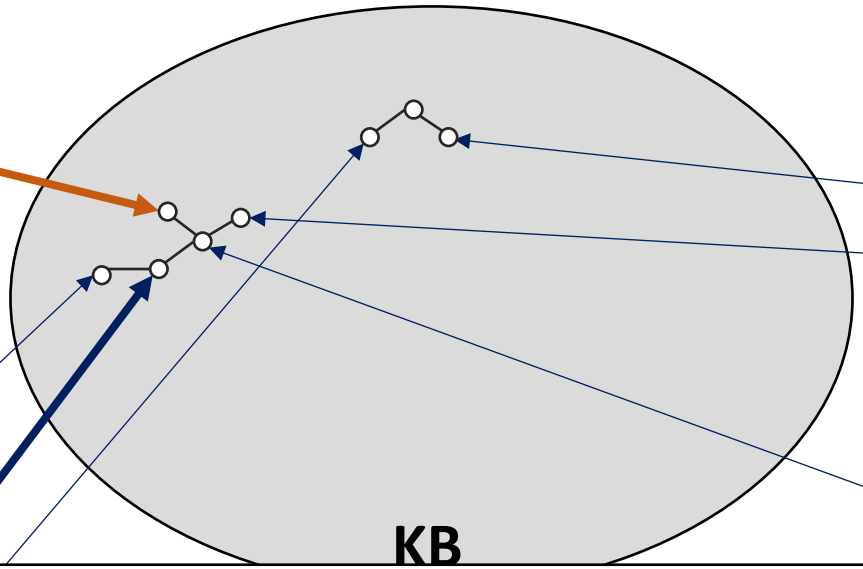
Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	goal scored by
4	film score
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	Croatia football team
16	Croatia basketball team
...	...



KB

conn(goal scored by**):**

0.50

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	France football team
...	...

2018 final

	KB item
1	2018 FIFA WC Final
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

KB-promixity

Who scored in the 2018 final between France and Croatia?

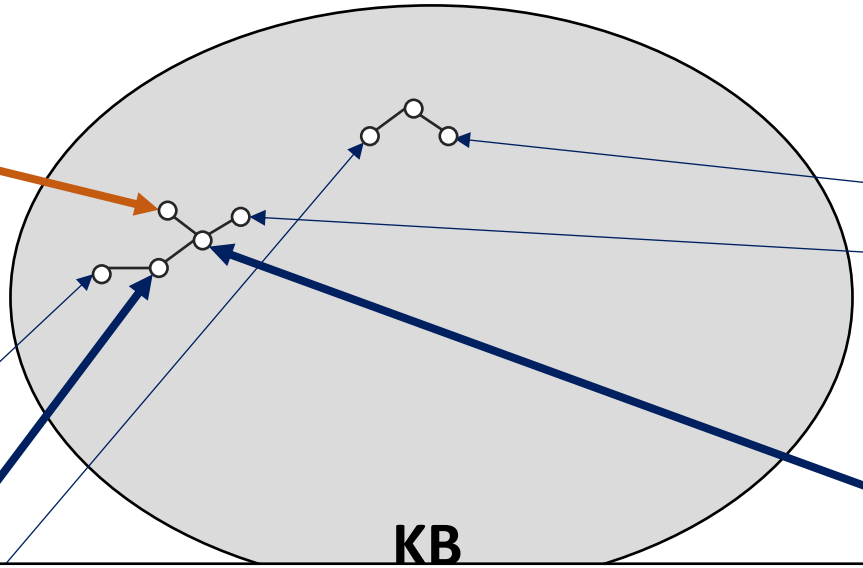
scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...

Symbolic space



KB

conn(goal scored by**):**

0.50 + 1.00

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

KB-promixity

Who scored in the 2018 final between France and Croatia?

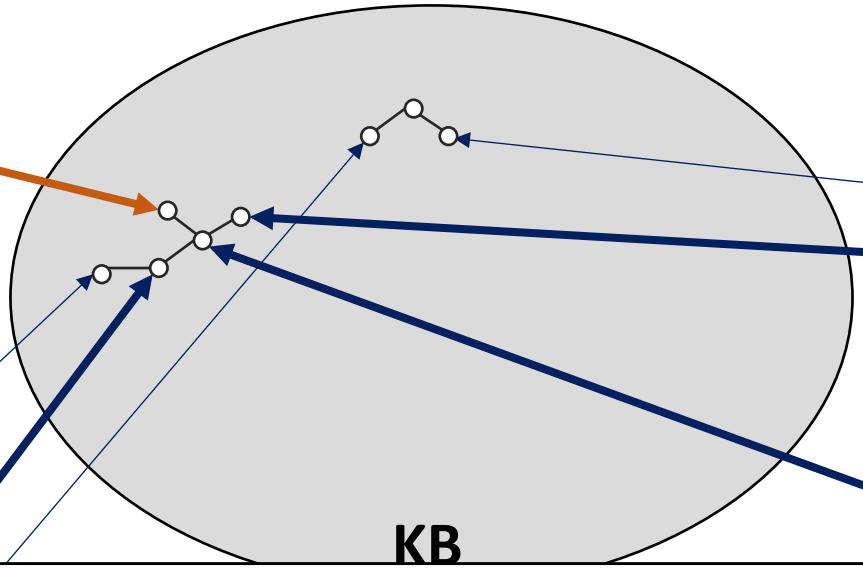
scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...

Symbolic space



KB

conn(goal scored by**):**

$\frac{1}{3} * (0.50 + 1.00 + 0.50) = 0.66$

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

KB-promixity

Who scored in the 2018 final between France and Croatia?

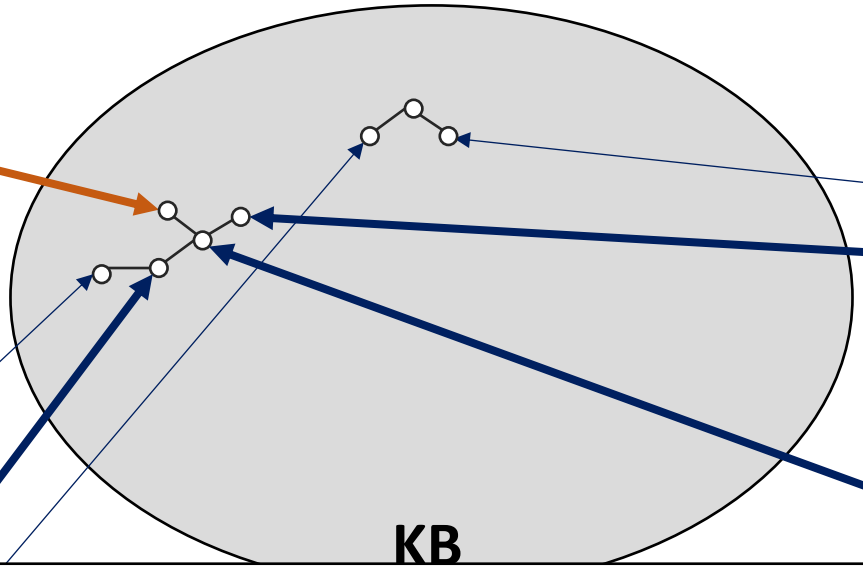
scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...

Symbolic space



KB

conn(goal scored by**):**

$\frac{1}{3} * (0.50 + 1.00 + 0.50) = 0.66$

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

Semantic coherence

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...

Latent space

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

Semantic coherence

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...

Latent space

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

Semantic coherence

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Latent space

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

coh(goal scored by**):**

0.96

Semantic coherence

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
...	...

Latent space

France

	KB item
1	France (state)
2	Kingdom of France
...	...
10	France basketball team
11	**France football team**
...	...

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
...	...
15	**Croatia football team**
16	Croatia basketball team
...	...

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
...	...

coh(goal scored by**):**

0.96 + 0.88

Semantic coherence

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	**goal scored by**
4	film score
…	…

Latent space

France

	KB item
1	France (state)
2	Kingdom of France
…	…
10	France basketball team
11	**France football team**
…	…

Croatia

	KB item
1	Croatia (state)
2	589 Croatia (asteroid)
…	…
15	**Croatia football team**
16	Croatia basketball team
…	…

2018 final

	KB item
1	**2018 FIFA WC Final**
2	2018 EuroLeague Final
3	Final Battle (event)
4	2018 FA Cup Final
…	…

coh(goal scored by**):**

$\frac{1}{3} * (0.96 + 0.88 + 0.89) = 0.91$

Question relevance

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	goal scored by
4	film score
...	...

Latent space

rel (goal scored by)**:**

0.76

Lexical matching

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	goal scored by
4	film score
...	...

match (goal scored by)**:**

$$\frac{1}{rank_{IR}+1} = \frac{1}{4} = 0.25$$

Candidate scoring

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	music score
2	no. of goals scored
3	goal scored by
4	film score
...	...

	KB item	score
1	goal scored by	0.58
2	no. of goals scored	0.52

$k = 2$

Rank

Aggregation

l_{12}	KB item	conn
1	goal scored by	0.66
2	no. of goals scored	0.50
3	film score	0.43
...	...	...

Pair-wise KB-proximity

l_{11}	KB item	coh
1	no. of goals scored	0.93
2	goal scored by	0.91
3	score (music)	0.81
...	...	...

Pair-wise semantic coherence

l_{13}	KB item	rel
1	goal scored by	0.76
2	score (music)	0.71
3	no. of goals scored	0.65
...	...	...

Question relevance

l_{14}	KB item	match
1	score (music)	0.50
2	no. of goals scored	0.33
3	goal scored by	0.25
...	...	...

Lexical matching

score = h_{coh} coh + h_{conn} conn + h_{rel} rel + h_{match} match

Search space retrieval

Who scored in the 2018 final between France and Croatia?

scored

	KB item
1	goal scored by
2	no. of goals scored

$k = 2$

2018 final

	KB item
1	2018 FIFA WC Final
2	2018 EuroLeague Final
3	2018 ChampionsLeague Final
4	2018 FA Cup Final

$k = 4$

France

	KB item
1	France football team
2	France basketball team
3	France (state)

$k = 3$

Croatia

	KB item
1	Croatia football team
2	Croatia basketball team
3	Croatia (state)

$k = 3$

⇒ Retrieve **KB-facts** with **disambiguated items**

⇒ **Novel KB-index** indexing full facts

⇒ Mechanism to keep **only salient facts**

- ⇒ E.g. US has millions of facts in the KB
- ⇒ If number of facts > parameter *p*, drop facts with item in object or qualifier-object position

CLOCQ KB-index

*"**Paul Pogba** scored a goal for the **France football team** in the **2018 FIFA WC Final**."*

⇒ **n-ary** fact with **qualifier** information
⇒ Represented in **reified** form via **dummy nodes**

⟨2018 FIFA WC Final, goal scored by, fact-id⟩
⟨fact-id, goal scored by, Paul Pogba⟩
⟨fact-id, for team, France football team⟩

WIKIDATA

⇒ **All facts** with Paul Pogba: expensive joining
⇒ Existing solutions optimize retrieval of **triples**

CLOCQ KB-index

⟨2018 FIFA WC Final, goal scored by, fact-id⟩
⟨fact-id, goal scored by, Paul Pogba⟩
⟨fact-id, for team, France football team⟩

$f = \langle \underbrace{\text{2018 FIFA WC Final, goal scored by, Paul Pogba}}_{\textbf{Main fact}}, \underbrace{\text{for team, France football team}}_{\textbf{Qualifier information}} \rangle$

⇒ Arbitrary length

KB-index:

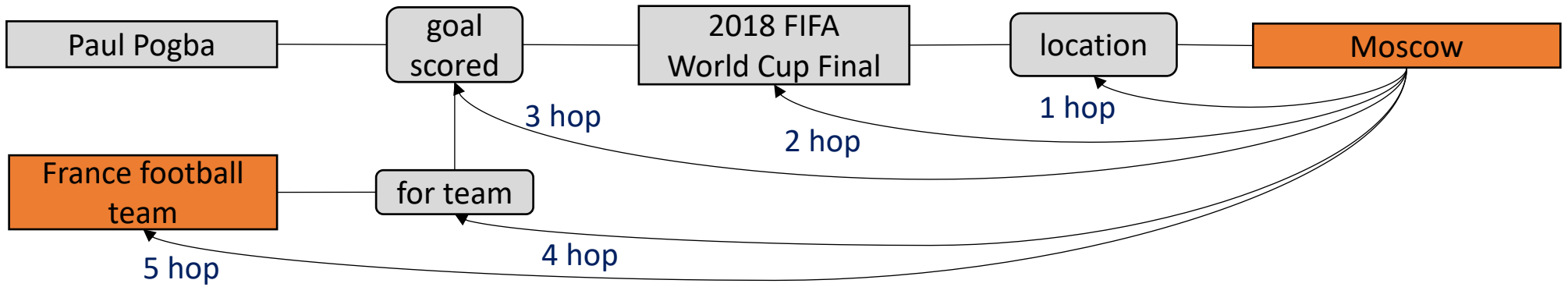
Index facts Paul Pogba ⟶ $\{ f \mid \text{Paul Pogba} \in f \}$

⇒ Efficient fact retrieval for entities

Index neighbors Paul Pogba ⟶ $\{ x \mid x \in f \wedge \text{Paul Pogba} \in f \}$

⇒ Efficient KB-distance computations via set operations

CLOCQ KB-index



⇒ **KB-distance:** d_k(*France football team*, *Moscow*) = **5**

⇒ **KB-distance:** d_k(*France football team*, *Moscow*) = **2**

CLOCQ KB-index

⟨2018 FIFA WC Final, goal scored by, fact-id⟩
⟨fact-id, goal scored by, Paul Pogba⟩
⟨fact-id, for team, France football team⟩

$f = \langle \underbrace{\text{2018 FIFA WC Final, goal scored by, Paul Pogba}}_{\textbf{Main fact}}, \underbrace{\text{for team, France football team}}_{\textbf{Qualifier information}} \rangle$

⇒ Arbitrary length

KB-index:

Index facts Paul Pogba ⟶ $\{ f \mid \text{Paul Pogba} \in f \}$

⇒ Efficient fact retrieval for entities

Index neighbors Paul Pogba ⟶ $\{ x \mid x \in f \wedge \text{Paul Pogba} \in f \}$

⇒ Efficient KB-distance computations via set operations

Search space retrieval

Who scored in the 2018 final between France and Croatia?

KB-facts with France football team

⟨**France football team**, type, football team⟩
⟨**France football team**, country, France⟩
⟨**France football team**, represents, French Football Federation⟩
⟨**France football team**, captain, Hugo Lloris⟩
⋮

63

⇒ as **subject**

Search space retrieval

Who scored in the 2018 final between France and Croatia?

KB-facts with France football team

⟨**France football team**, type, football team⟩
⟨**France football team**, country, France⟩
⟨**France football team**, represents, French Football Federation⟩
⟨**France football team**, captain, Hugo Lloris⟩
⋮ **63**

⟨FIFA WC 2018, winner, **France football team**⟩
⟨FIFA WC 1998, winner, **France football team**⟩
⟨France v Albania, winner, **France football team**⟩ ⟩
⟨Antoine Griezmann, member of sports team, **France football team**, start, 2014⟩
⟨Zinedine Zidane, member of sports team, **France football team**, start, 1994, end, 2006⟩
⋮ **1,044**

⇒ as **object**

Search space retrieval

Who scored in the 2018 final between France and Croatia?

KB-facts with France football team

⟨**France football team**, type, football team⟩
⟨**France football team**, country, France⟩
⟨**France football team**, represents, French Football Federation⟩
⟨**France football team**, captain, Hugo Lloris⟩
⋮ — **63**

⟨FIFA WC 2018, winner, **France football team**⟩
⟨FIFA WC 1998, winner, **France football team**⟩
⟨France v Albania, winner, **France football team**⟩)
⟨Antoine Griezmann, member of sports team, **France football team**, start, 2014⟩
⟨Zinedine Zidane, member of sports team, **France football team**, start, 1994, end, 2006⟩
⋮ — **1,044**

⟨1958 FIFA WC, statistical leader, Just Fontaine, criterion, goals scored, for team, **France football team**⟩
⟨2003 Confed Cup, statistical leader, Thierry Henry, criterion, goals scored, for team, **France football team**⟩
⟨Benjamin Mendy, participant in, 2018 FIFA WC, #matches played, 1, for team, **France football team**⟩
⟨Antoine Griezmann, participant in, 2018 FIFA WC, #matches played, 7, for team, **France football team**⟩
⟨Antoine Griezmann, participant in, 2014 FIFA WC, #matches played, 5, for team, **France football team**⟩
⋮ — **147**

⇒ as **qualifier-object**

Search space retrieval

Who scored in the 2018 final between France and Croatia?

KB-facts with France football team

⟨**France football team**, type, football team⟩
⟨**France football team**, country, France⟩
⟨**France football team**, represents, French Football Federation⟩
⟨**France football team**, captain, Hugo Lloris⟩
⋮

63

⟨FIFA WC 2018, winner, **France football team**⟩
⟨FIFA WC 1998, winner, **France football team**⟩
⟨France v Albania, winner, **France football team**⟩)
⟨Antoine Griezmann, member of sports team, **France football team**, start, 2014⟩
⟨Zinedine Zidane, member of sports team, **France football team**, start, 1994, end, 2006⟩
⋮

1,044

⟨1958 FIFA WC, statistical leader, Just Fontaine, criterion, goals scored, for team, **France football team**⟩
⟨2003 Confed Cup, statistical leader, Thierry Henry, criterion, goals scored, for team, **France football team**⟩
⟨Benjamin Mendy, participant in, 2018 FIFA WC, #matches played, 1, for team, **France football team**⟩
⟨Antoine Griezmann, participant in, 2018 FIFA WC, #matches played, 7, for team, **France football team**⟩
⟨Antoine Griezmann, participant in, 2014 FIFA WC, #matches played, 5, for team, **France football team**⟩
⋮

147

> 1k

⇒ Keep only **salient facts**

Intuition:

⇒ If **too many facts** with item as object or qualifier-object

⇒ **Drop them**

⇒ Too many: **more than *p***

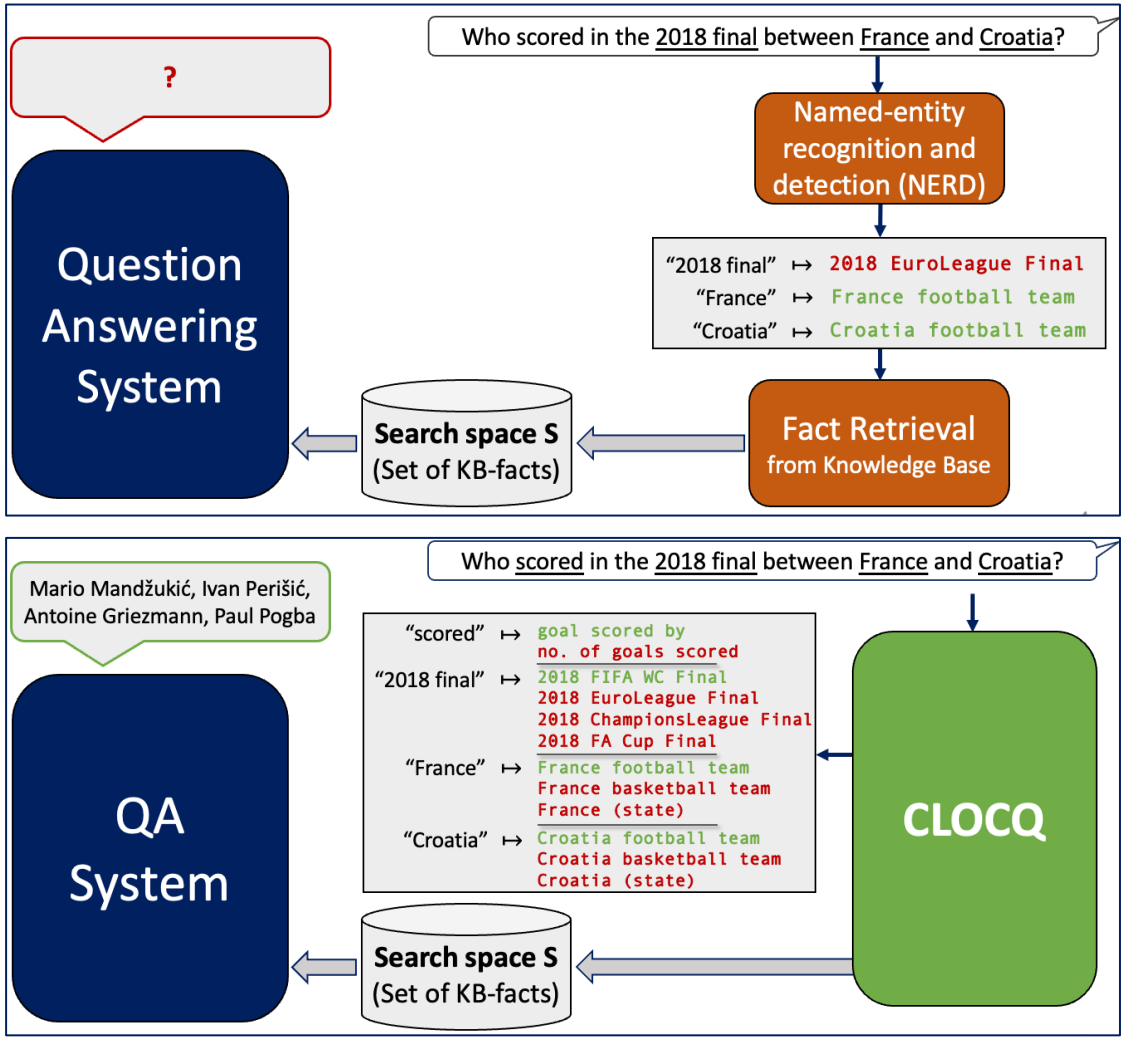
⇒ Experiments with **different *p***

Baselines

Baselines:
NERD + KB-fact retrieval

NERD:
TagME [Ferragina and Scaiella, CIKM 2010]
ELQ [Li et al., EMNLP 2020]
AIDA [Hoffart et al., EMNLP 2011]
REL [van Hulst et al., SIGIR 2020]
EARL [Dubey et al., ISWC 2018]

KB-fact retrieval:
HDT [Fernández et al., JWS 2013]
SPARQL queries



Experimental setup

Knowledge Base

⇒ **Wikidata** (Initial search space has 152 million items)

Benchmarks

⇒ **LC-QuAD2.0** (10,000 complex questions) [Dubey et al., ISWC 2019]

⇒ **ConvQuestions** (1,569 full questions) [Christmann et al., CIKM 2019]

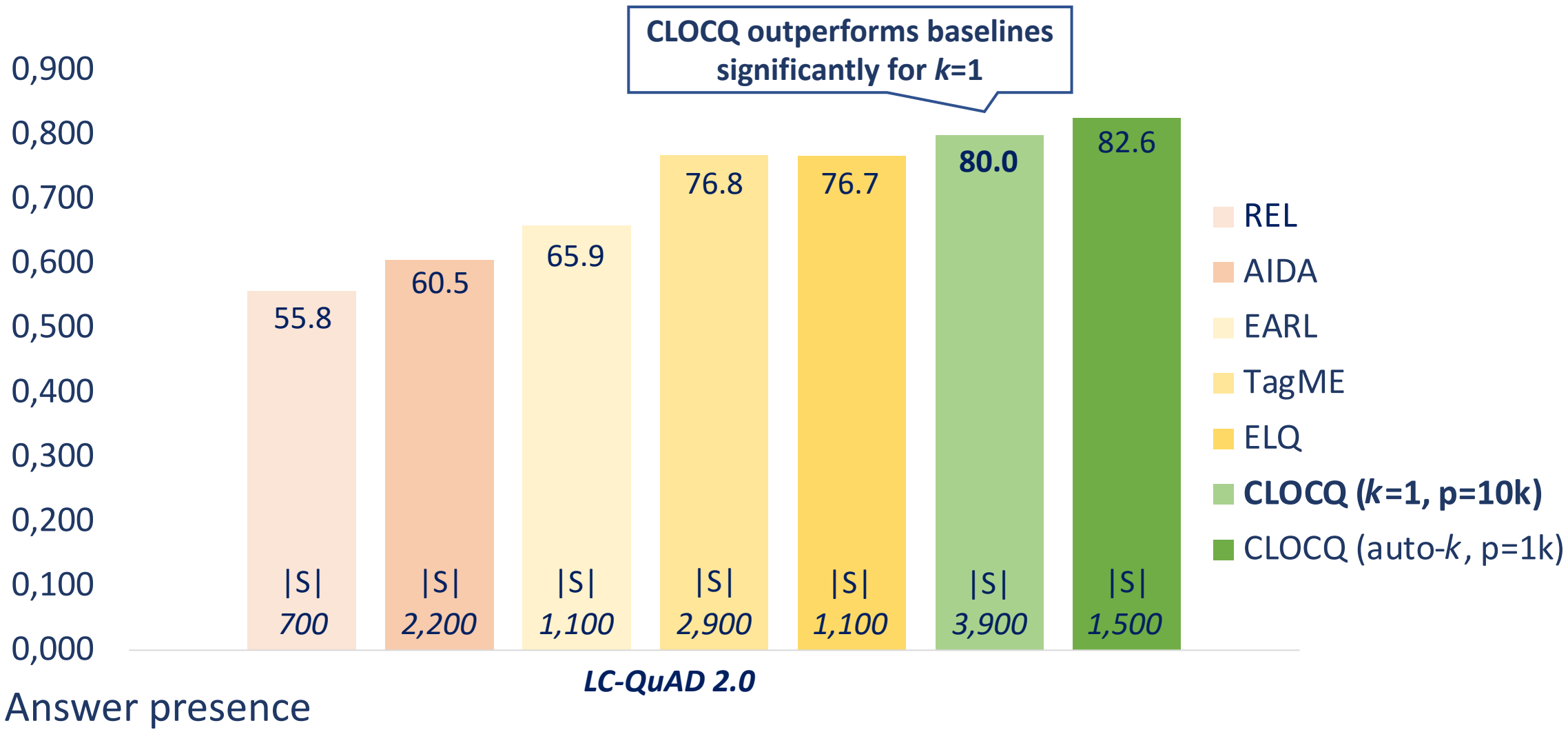
Metrics

⇒ **Answer presence** (answer in search space?)

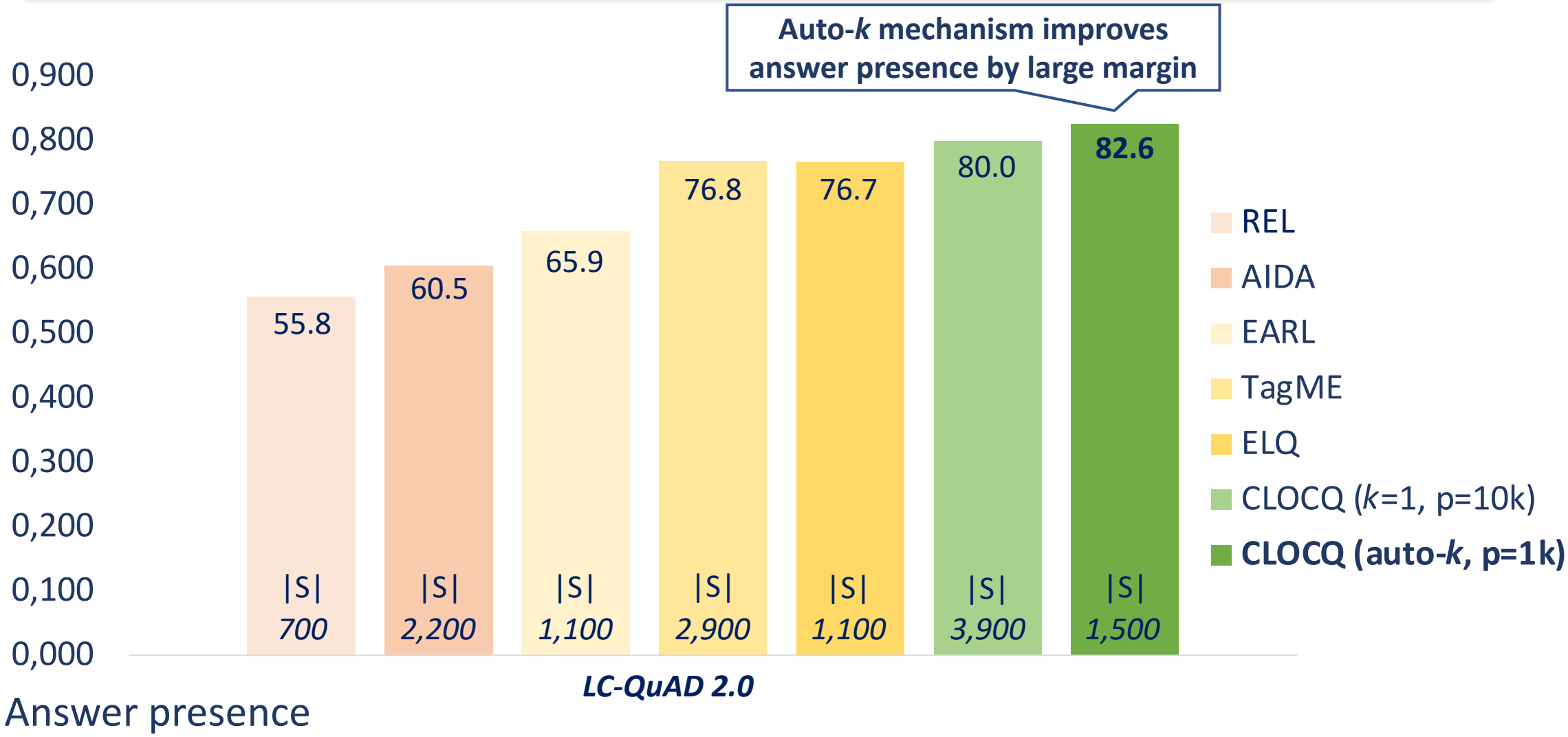
⇒ **Search space size |S|** (number of answer candidates)

⇒ **Runtime** (user question → search space)

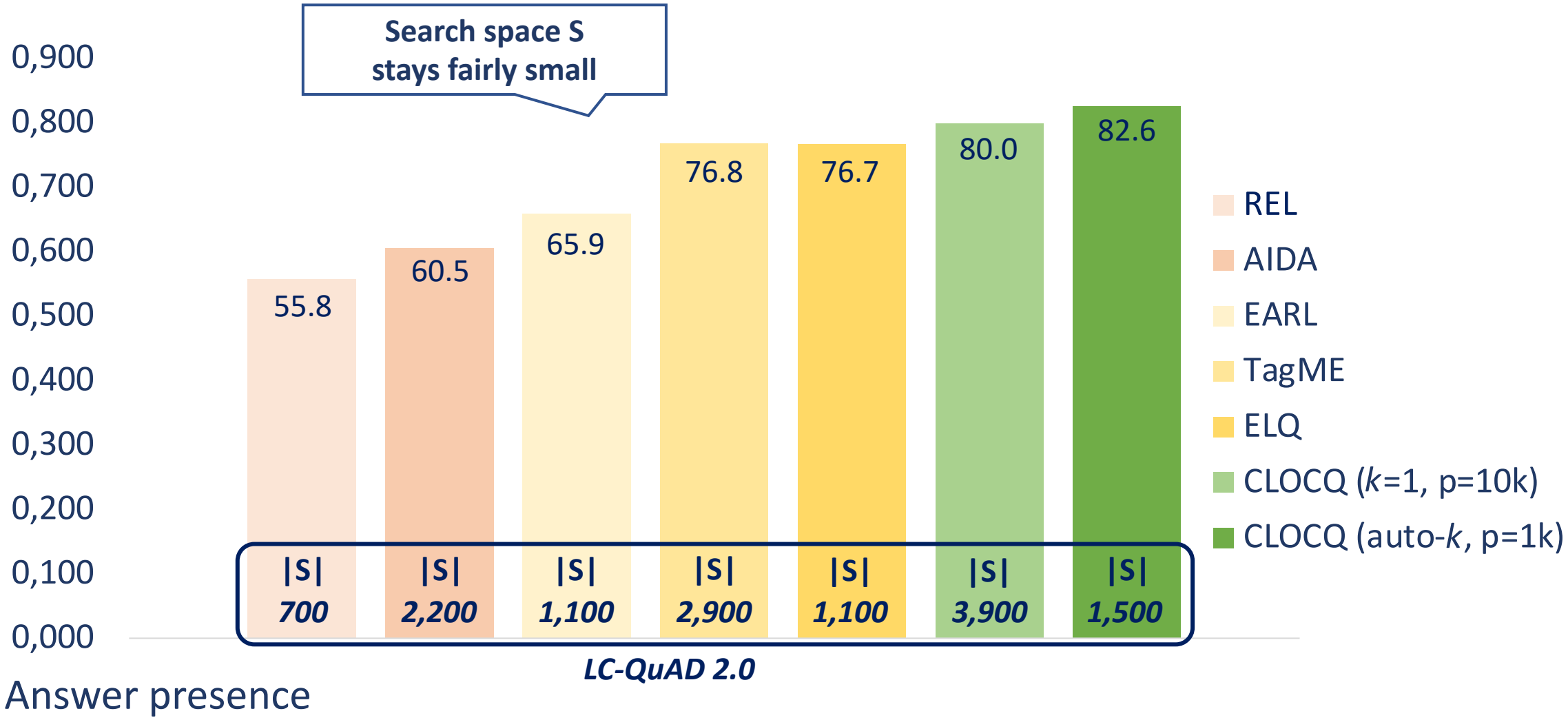
Answer presence: LC-QuAD 2.0



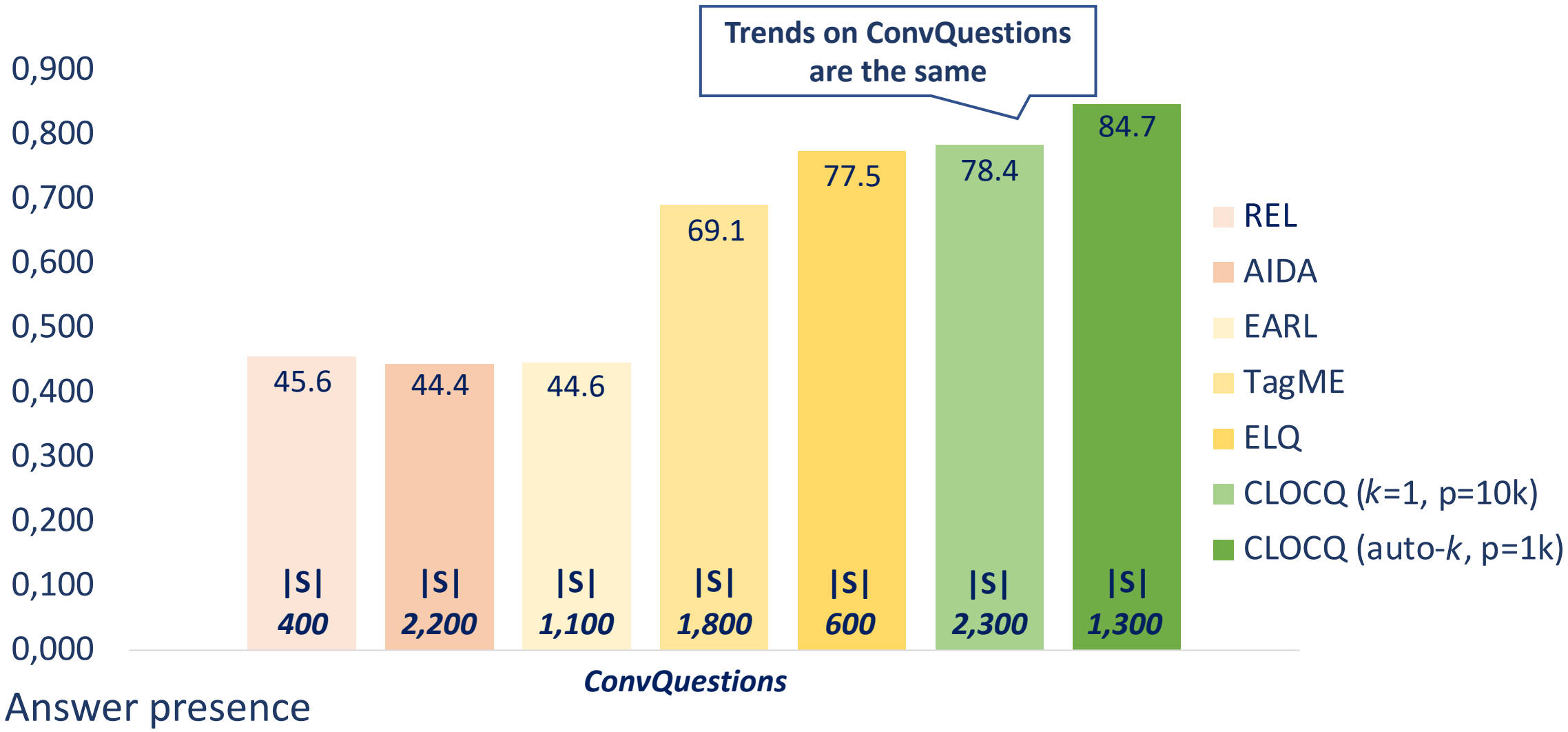
Answer presence: LC-QuAD 2.0



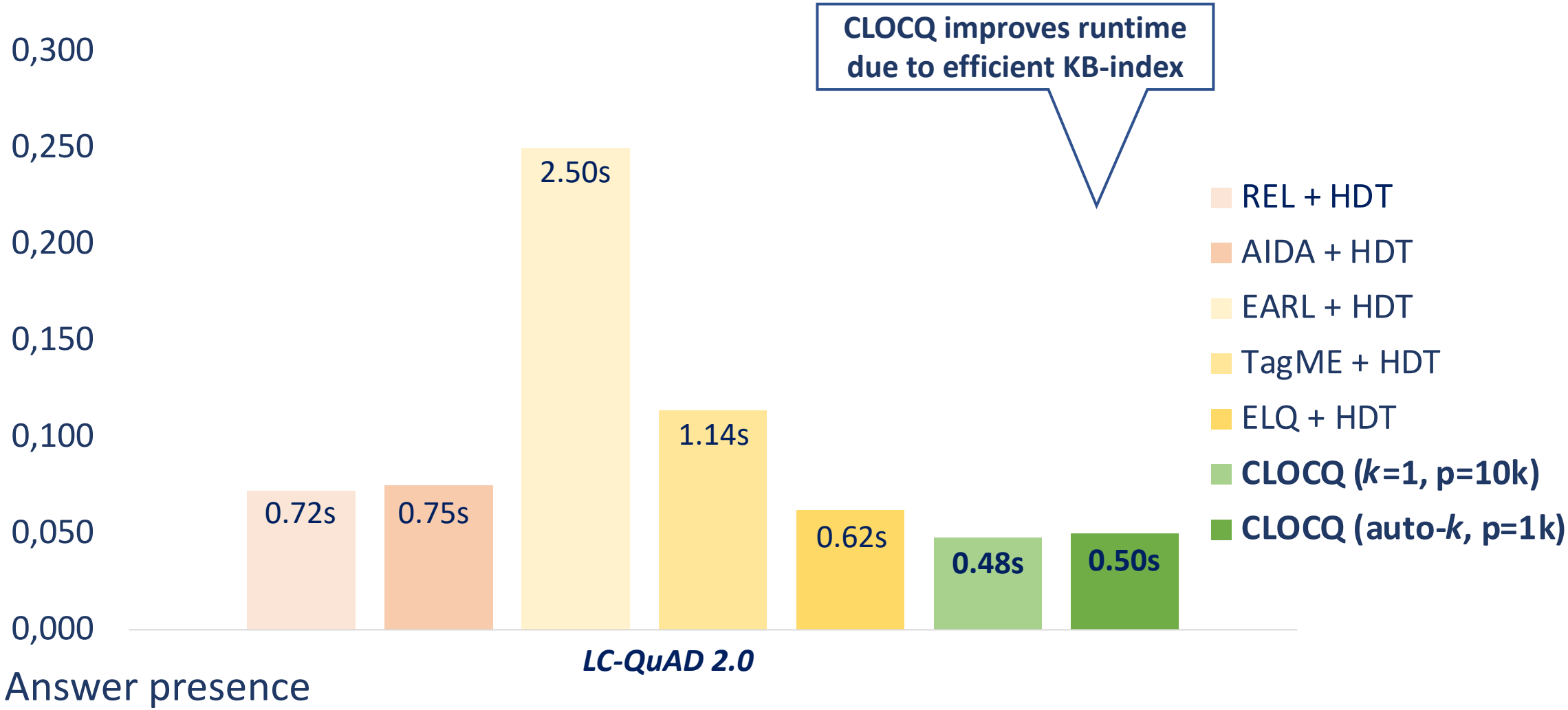
Answer presence: LC-QuAD 2.0



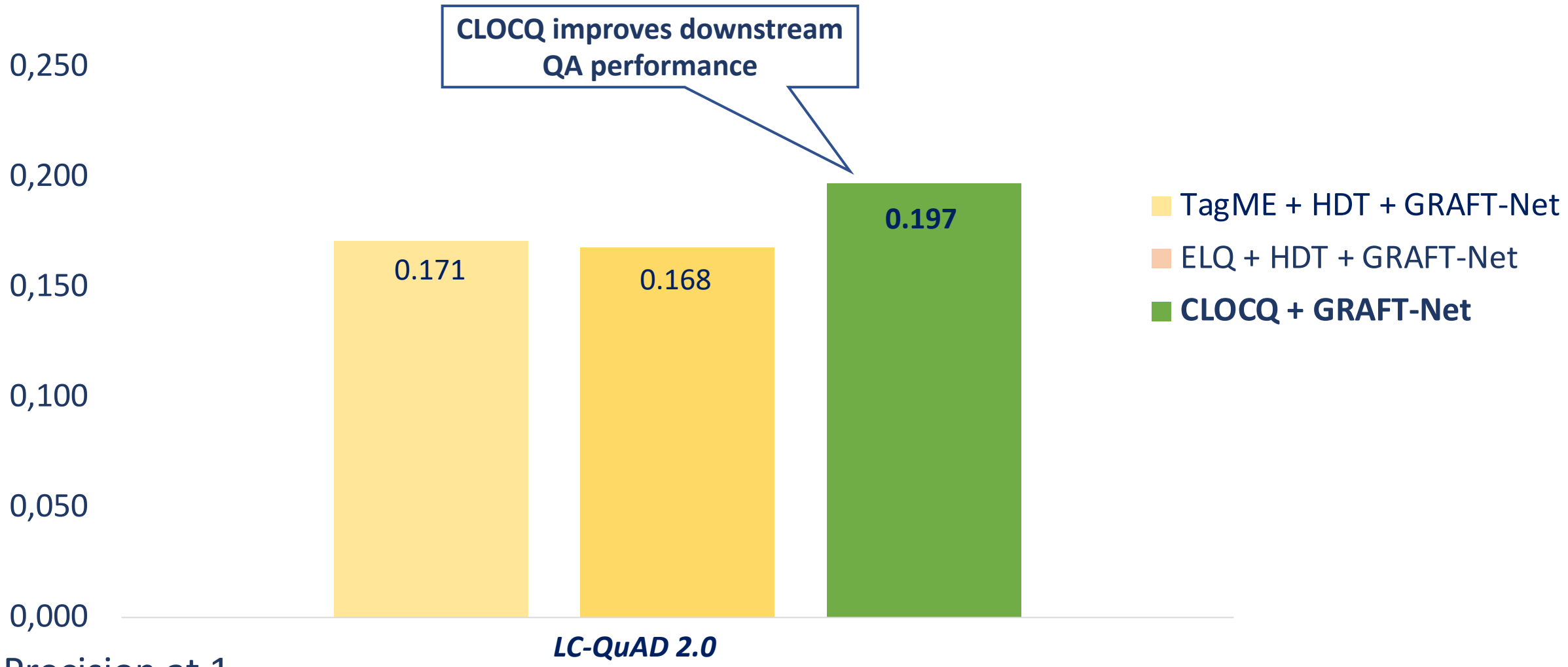
Answer presence: ConvQuestions



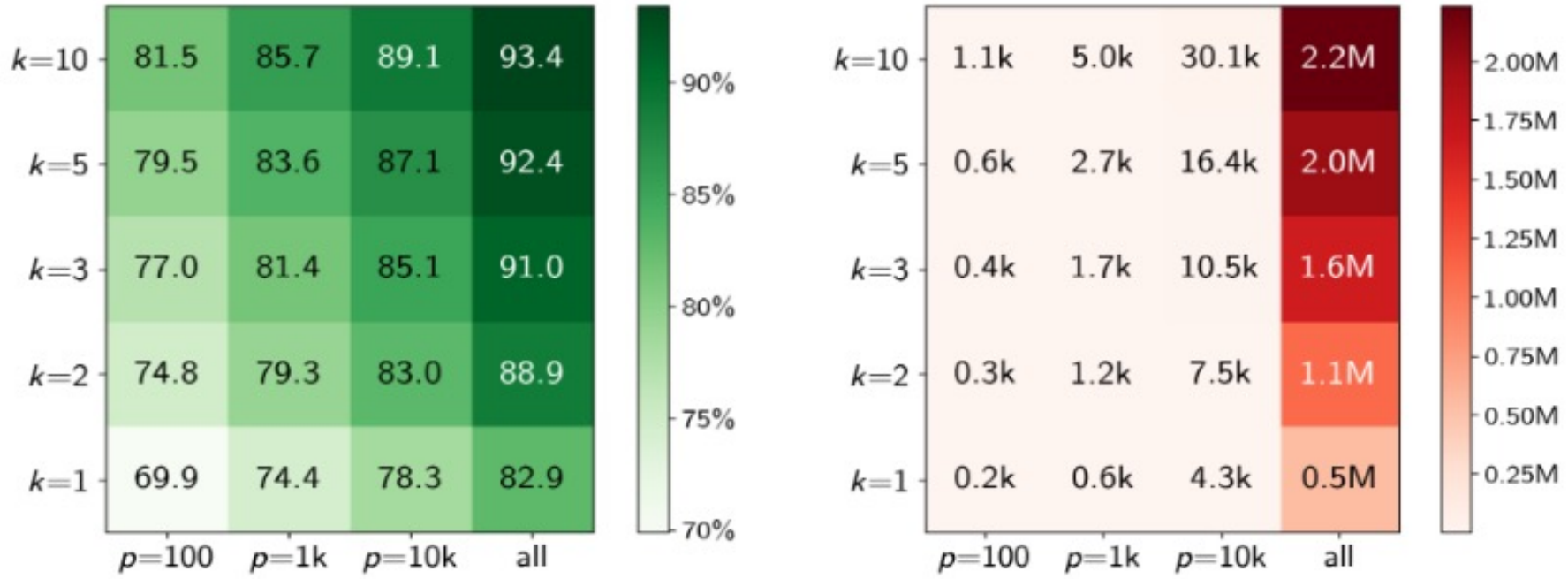
Runtime


CLOCQ improves runtime due to efficient KB-index
0,300
0,250
0,200
0,150
0,100
0,050
0,000
0.72s
0.75s
2.50s
1.14s
0.62s
0.48s
0.50s
REL + HDT
AIDA + HDT
EARL + HDT
TagME + HDT
ELQ + HDT
CLOCQ (k=1, p=10k)
CLOCQ (auto-k, p=1k)
LC-QuAD 2.0
Answer presence

End-to-end QA: LC-QuAD 2.0

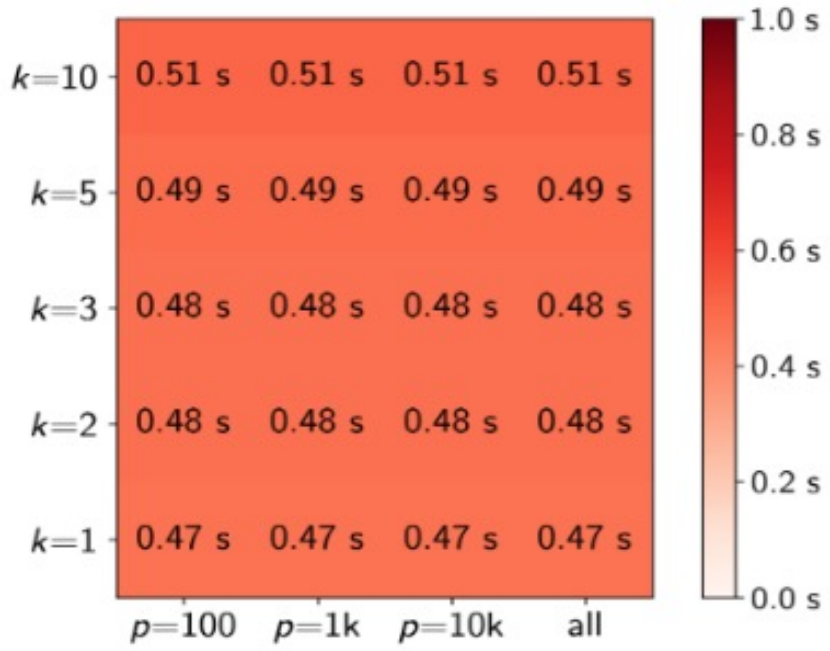


Trade-off: increasing *k* and *p*



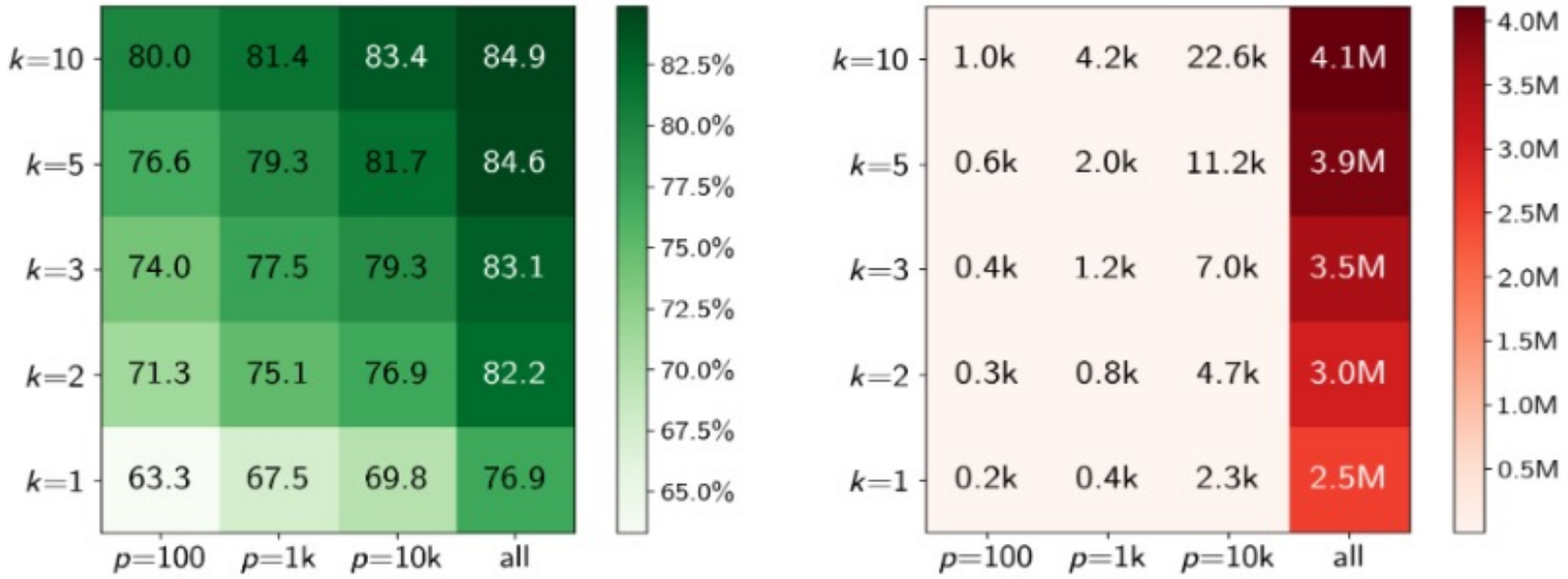
LC-QuAD 2.0

Trade-off: increasing *k* and *p*



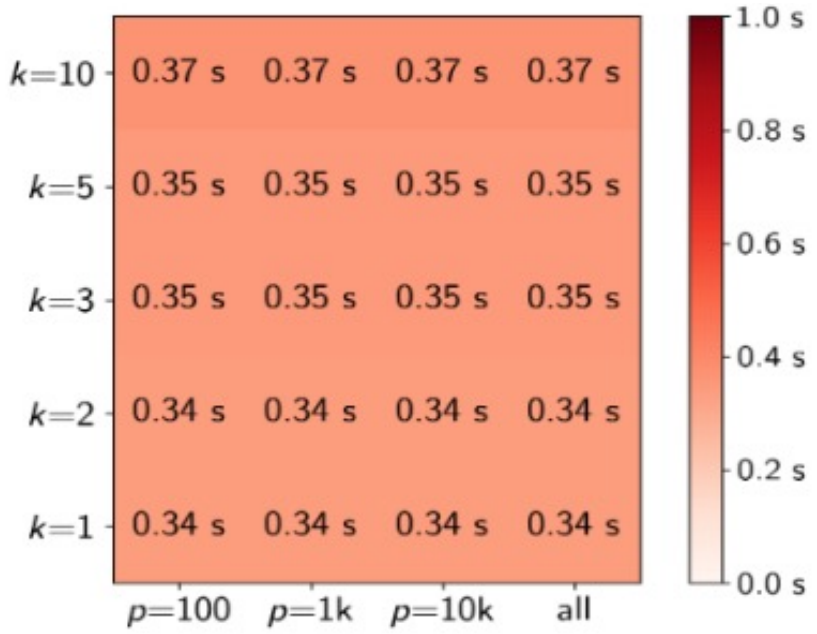
LC-QuAD 2.0

Trade-off: increasing *k* and *p*



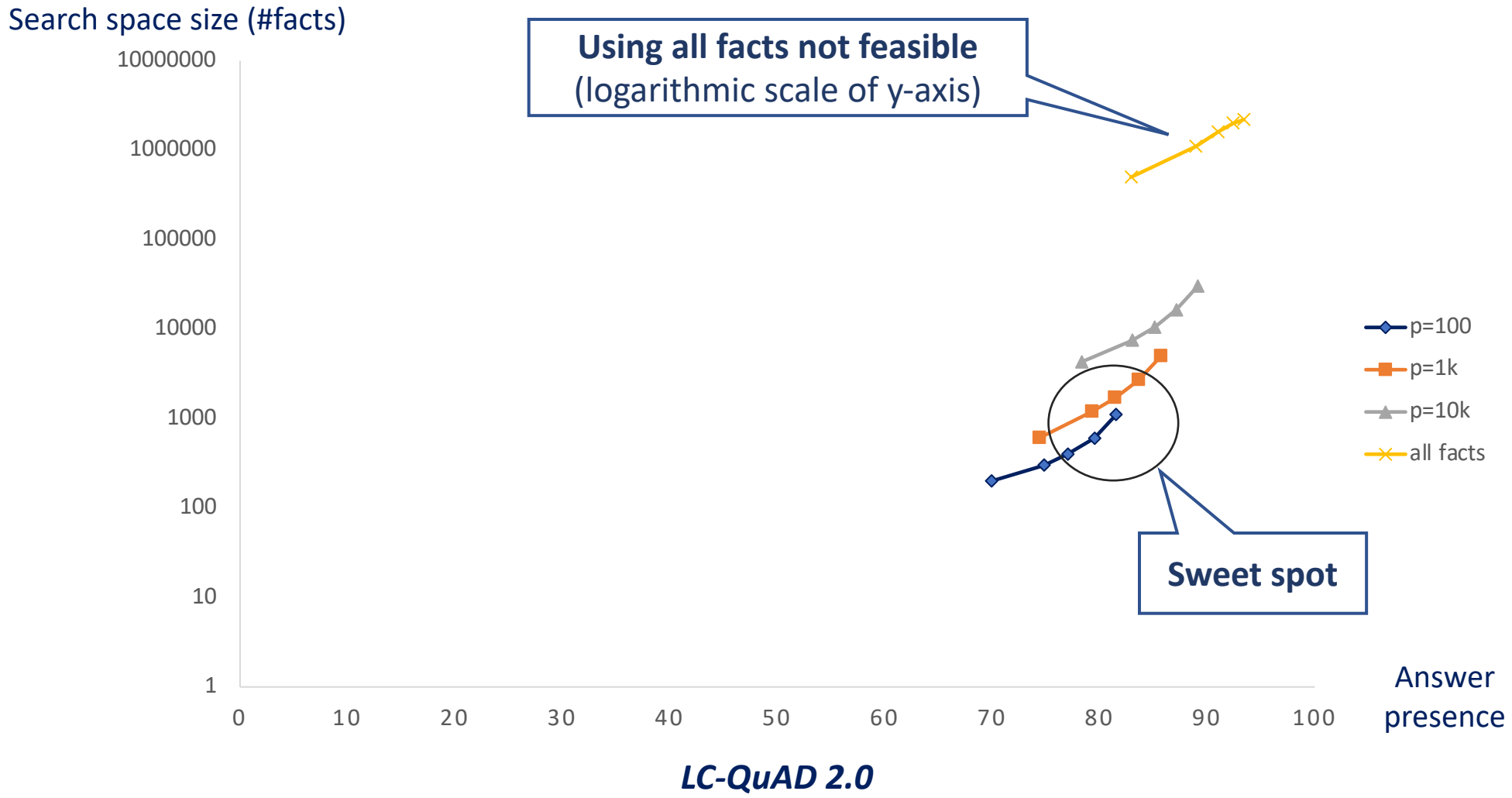
ConvQuestions

Trade-off: increasing *k* and *p*



ConvQuestions

Trade-off: answer presence – #facts



LC-QuAD 2.0

Conclusion

Identify **search space reduction** as a **critical task** in KB-QA pipeline

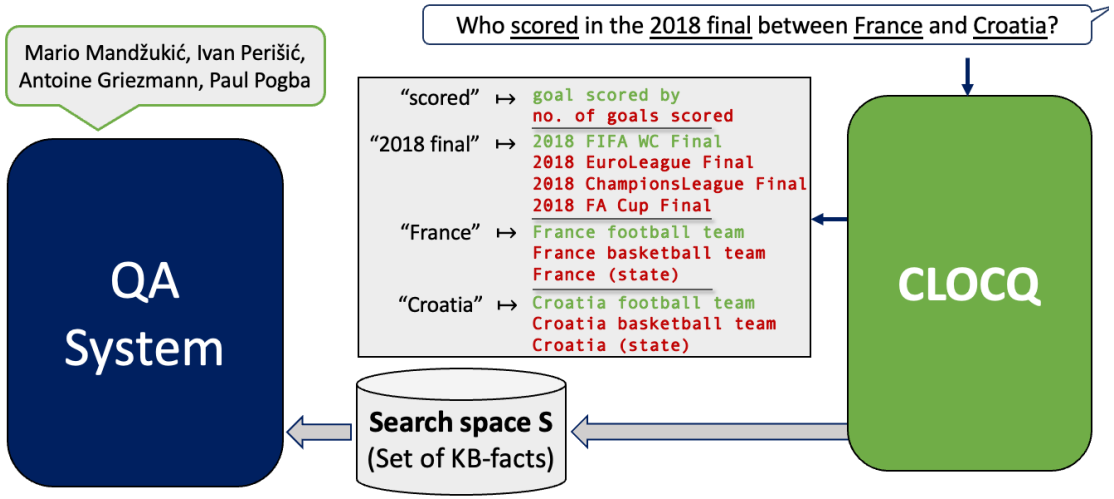
CLOCQ: end-to-end search space reduction, going beyond NERD

⇒ Disambiguate **all KB-items** (entities, types, concepts, predicates)

⇒ Top-***k*** **disambiguations**

⇒ ***k*** set automatically for each **individual** question word

⇒ Integrate **KB-aware signal** in scoring



New to KB-QA?

CLOCQ allows simple access to KB to get started

⇒ Search space reduction for given question

⇒ Retrieve all facts of KB-item from KB

clocq.mpi-inf.mpg.de

Thank you!