TempQuestions: A Benchmark for Temporal Question Answering

Zhen Jia[1], Abdalghani Abujabal[2], Rishiraj Saha Roy[2], Jannik Strötgen[2], Gerhard Weikum[2]

[1]SW Jiaotong University, China   [2]Max Planck Institute for Informatics, Germany

HQA2018, 23 April 2018, Lyon, France

Why a new benchmark?

★ Answering complex questions current focus of community

★ Several types of (factoid) complex questions

★ Temporal questions need special methods!!

"who won best supporting actor when alfred junge won best art direction?"

"when was the world wide web conference held in lyon before 2018?"

★ Fair comparison requires benchmark
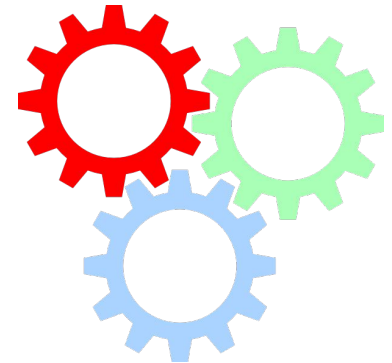
It's About Time

A clear definition

★ Benchmarking necessitates crisp boundaries

★ Not trivially "something about time"

★ *Our definition:* **"A temporal question is any question, which contains a temporal expression, a temporal signal, or whose answer is of temporal nature."**

★ Based on TimeML concepts: http://www.timeml.org/

It's About Time
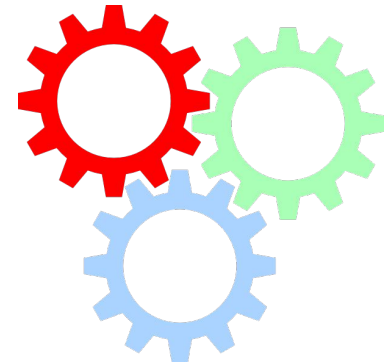
Concepts: Temporal Expressions

★ TimeML tag TIMEX3

  ○ Dates, times, durations, sets

  ○ *23 april 2018, 4 pm, six months, once a year*

  ○ Explicit and implicit
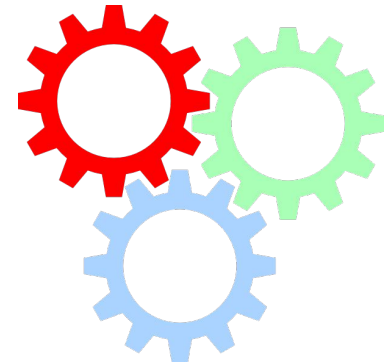
  ○ *summer olympics 2016, trump's presidency*

Concepts: Temporal Signals

★ TimeML tag SIGNAL

★ 13 temporal relations from Allen's interval algebra (Allen 1983)

for temporal reasoning - *starts, meets, overlaps, before, during, …*

★ Ordinal expressions

★ Examples: *before, after, following, during, while,*

first, second, last, …

Concepts: Answers of Temporal Nature

★   Answers are explicit or implicit temporal expressions

★   Usually indicated by start patterns

★   Examples: *when, in which year, ...*

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Creating TempQuestions...

★ QA benchmarks contain small fractions of temporal questions

★ Systems ignoring temporal questions achieve satisfactory performance, effect lost in aggregation

★ Extracted temporal questions from 3 popular KB-QA benchmarks:

  ○ **Free917** (Cai and Yates 2013)

  ○ **WebQuestions** (Berant and Liang 2013)

  ○ **ComplexQuestions** (Bao et al. 2016)

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Detecting Temporal Questions..

★ **Temporal expressions:** HeidelTime (Strötgen and Gertz 2015) and SUTime (Chang and Manning 2012)

★ **Events:** Dictionary from Freebase (*world war II, super bowl 2013*)

★ **Temporal signals:** Dictionary suggested by Setzer (2002)

★ **Ordinal words:** Stanford CoreNLP (Manning et al. 2014)

★ **Start patterns:** *when, what date, in which year, …*

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Detecting Temporal Questions..

★ Questions with any of these tags added to pool

★ 1,541 questions in initial pool

★ Focused on recall of temporal questions

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Cleaning Temporal Questions

★   False positives removed

who does nicolas cage play in a christmas carol?

which countries have german as the first language?

who was the director of 2012?
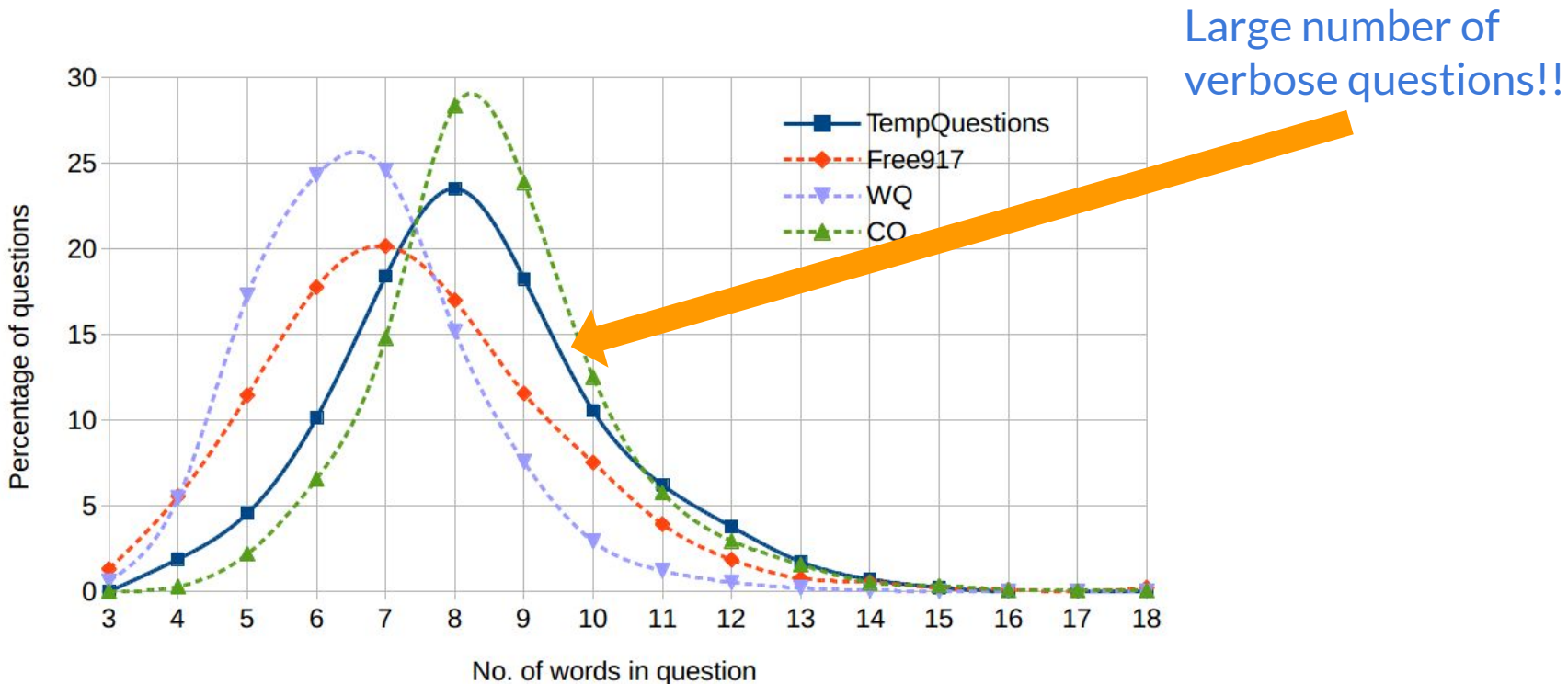
★   Answers sets checked for accuracy and redundancy

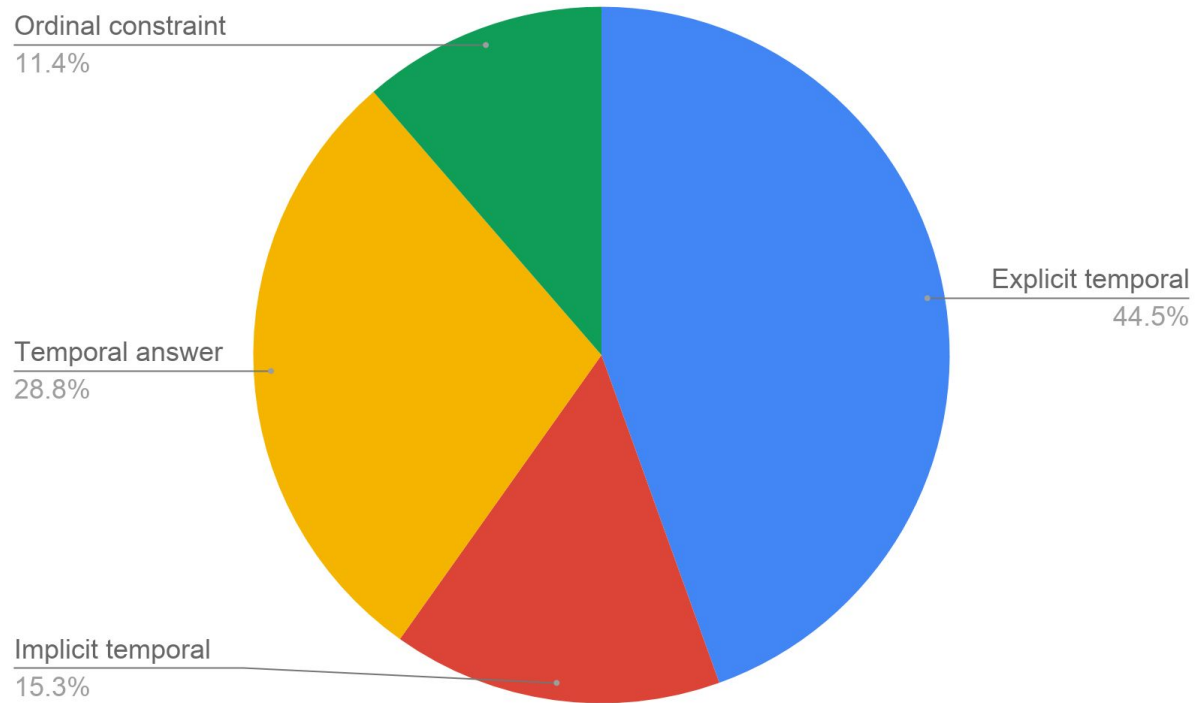*who won the grammy for best new artist in 2014? [Ryan Lewis][Sam Smith]*

who did libya gain independence from in 1951? [Italy, Independence from Italy]

★   1,271 questions in final benchmark

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Question Length Distribution

Large number of verbose questions!!

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Casting into Question Categories...

Ordinal constraint
11.4%

Explicit temporal
44.5%

Temporal answer
28.8%

Implicit temporal
15.3%

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

12

Casting into Question Categories...

★ Questions with **explicit temporal expressions**

who won the state of texas in 2008?

what kind of government does iran have after 1979?

who is the american author who lived in japan during the 1800's?

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

13

Casting into Question Categories...

★ Questions with **implicit temporal expressions**

who did Michael Jordan play for when he played baseball?

who was the president after jfk died?

which movie won best cinematography when sam mendes won best director?

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Casting into Question Categories...

★ Questions with **temporal answers**

when was the united nations founded?

what years did the knicks win the championship?

what year did baltimore ravens win super bowl?

Casting into Question Categories...

★   Questions with **ordinal constraints**

what is the name of dawn french's first novel?

who was andy williams second wife?

when is the last time the raiders won a championship?

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Challenges in TempQuestions..

★ Questions with **multiple entities (205)**

what did france lose to the british in the treaty of paris in 1763?

what country did united states purchase alaska from in 1867 for $7.2 million dollars?

when was the last time the oakland raiders won the super bowl?

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Challenges in TempQuestions..

★ Questions with **multiple relations (145)**

who won best supporting actor when alfred junge won best art direction?

what book was written by george orwell and published in 1945?

which film starring vyacheslav tikhonov was released in 1994?

Challenges in TempQuestions..
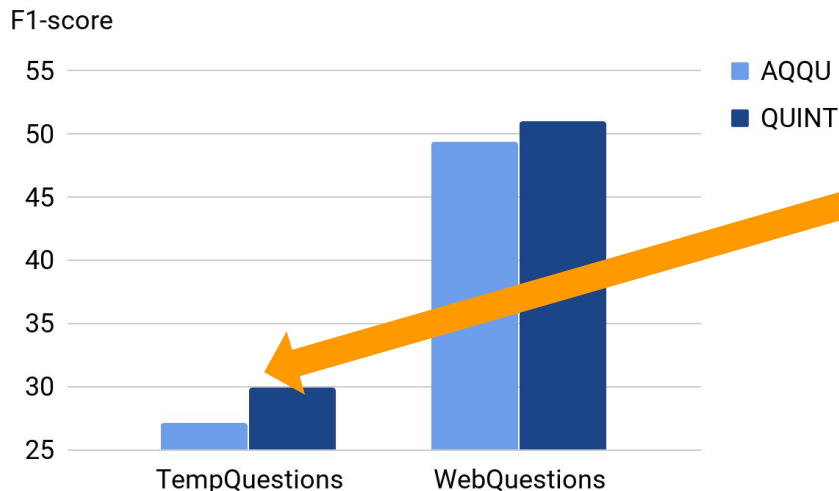
★ Questions with **signal words**

- **Before (49):** *who lived in america before europeans arrived?*

- **After (28):** *what team did michael jordan play for after the bulls?*

- **Overlap (435):** *who was the leader of the us during wwii?*

- **Ordinal (156):** *who was the first coach of the buccaneers?*

Performance of state-of-the-art systems

★ **AQQU** (Bast and Haussmann 2015) and **QUINT** (Abujabal et al. 2017)

★ Compared on **TempQuestions** and **WebQuestions**

Big room for improvement!!

Particularly bad on implicit questions!! (~20%)

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

Summarizing...

★ **TempQuestions**, a new benchmark for temporal QA with

   1,271 QA pairs + clear **definition** of temporal questions

★ Answering requires question **decomposition and reasoning**

★ Questions have useful **markup tags**

★ Available at: **http://qa.mpi-inf.mpg.de/TempQuestions.zip**

★ Suitable for both **KB-QA and IR-/text-QA**

More on QA from us @MPI for Informatics!!

Never-Ending Learning for Open-Domain Question Answering over Knowledge Bases

Wed 25 Apr, 14:30, Web Content Analysis II, Salon Tête d'Or

TempQuestions: A Benchmark for Temporal Question Answering, HQA2018

21