CLOCQ can be used for efficient retrieval of relevant KB-facts for any query

**SEARCH SPACE REDUCTION: OVERLOOKED PROBLEM IN KB-QA**
- Curated knowledge bases (KB) have billions of facts, with millions of entities and thousands of predicates
- Question answering (QA) system typically requires input to be sufficiently small
- Therefore, QA system not directly applicable to full KB
- Typical KB-QA pipeline first reduces the search space to a few thousands of facts, that are relevant to the question and likely to contain the answer(s)

**STANDARD NED: ONE SIZE FITS ALL**
- Top-1 NED too restrictive for QA (single error can cause failure of full QA pipeline)
- Top-k NED can easily include noise, blowing up the search space
- Important cues within the question (e.g. scored in example on the right) are not considered, given that most NED methods disambiguate entity mentions only

**APPROACH: GOING BEYOND NED**
- Adaptive top-k: choose appropriate k for each question word individually
- Disambiguate all KB-items (entities, types, concepts, predicates)
- Consider proximity of disambiguated KB-items within graph underlying the KB
- Establish a novel fact-centric KB-index for more efficient access to the KB

**PIPELINE: SEARCH SPACE REDUCTION FOR COMPLEX KB-QA**
1. KB-item retrieval: Retrieve d KB-item candidates per question word
2. Auto-k: Choose k automatically for each question word based on its ambiguity
3. Candidate scoring: Identify top-k candidates using signals for pair-wise KB-proximity, pair-wise semantic coherence, question relevance, lexical matching
4. Search space retrieval: Retrieve facts for disambiguated KB-items from knowledge base using our fact-centric KB-index: consider only salient facts, pruning potentially noisier facts based on a parameter p (e.g. facts with US as object or qualifier-object)

**CLOCQ makes use of a fact-centric KB-index, which stores n-nary facts in a way that allows for efficient retrieval.**

**RESULTS ON LC-QUAD 2.0**
- **CLOCQ**: 82.6
- **CLOCQ (r=3)**: 80.0
- **TagME+HDT**: 76.8
- **ELQ+HDT**: 76.7
- **EARL**: 60.5
- **AIA+D**: 60.5
- **REL+HDT**: 55.8

**LC-Quad 2.0**
- Question: “Who is the composer of All We Know?”
  
  CLOCQ: Composer: film score composer, AllWeKnow=[(All We Know (Paramore))].
  HDT: Composer=All We Know.

**SIC**

**Contact**: pchristm@mpi-inf.mpg.de
**More information**: clocq.mpi-inf.mpg.de