Multi-Mention Learning for Reading Comprehension using Neural Cascades

Carnegie Mellon University
Language Technologies Institute

Swabha Swayamdipta
Carnegie Mellon University

Ankur Parikh   Tom Kwiatkowski
Google Research

Research at Google

Reading Comprehension

Which famous novelist also wrote under the name Richard Bachman?
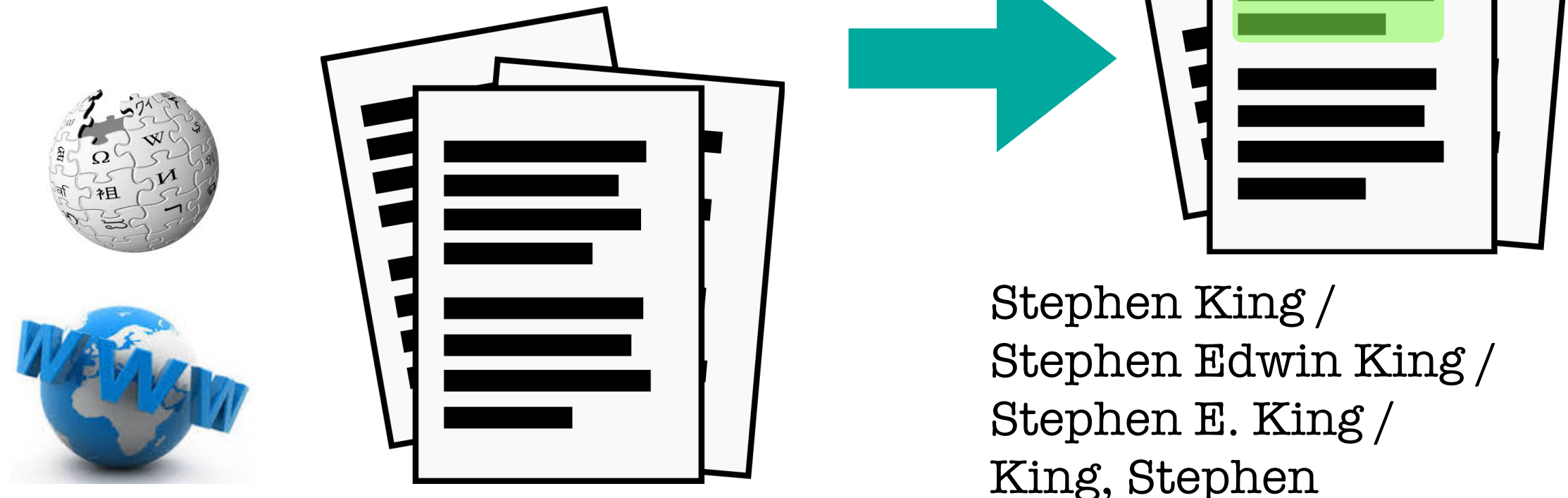
Stephen King /
Stephen Edwin King /
Stephen E. King /
King, Stephen

Distant Supervision

- TriviaQA (Joshi et. al., 2017) with questions and answers taken from trivia websites.
- Wikipedia and web pages provided as sources of evidence, *independent* of the QA pairs.
- Documents contain the answer strings, but are not guaranteed to contain any *evidence* to answer the questions.

Question	Which US artist married Lee Krasner in 1945?
Answer	Jackson Pollock; Pollock; Pollock, Jackson
Document	Wikipedia entry for Lee Krasner (Excerpts shown)

Start	Passage
952	She lost interest in their usage of hard-edge geometric style after her relationship with Pollock began.
3364	Pollock's influence
3366	Although many people believe that Krasner stopped working in the 1940s in order to nurture Pollock's home life and career, she never stopped creating art.
4084	Relationship with Jackson Pollock
4089	Lee Krasner and Jackson Pollock established a relationship with one another in 1942 after they both exhibited at the McMillen Gallery. In the summer of 1945, they got married in a
4560	While she married Pollock in a church, Krasner continued to identify herself as Jewish but decided to not practice the religion.

Challenges

- Read multiple documents in their entirety.
- Answers occurrences, though repeated, may not contain adequate evidence to answer the question.

Longer Document Context

- Longer the context considered, better the chances of correctly answering a question.
- Tradeoff : efficiency.
- **Simple feed-forward nets.**

Multiple Mentions of Answers

- Not each mention of the answer is relevant or sufficient to answer the question.
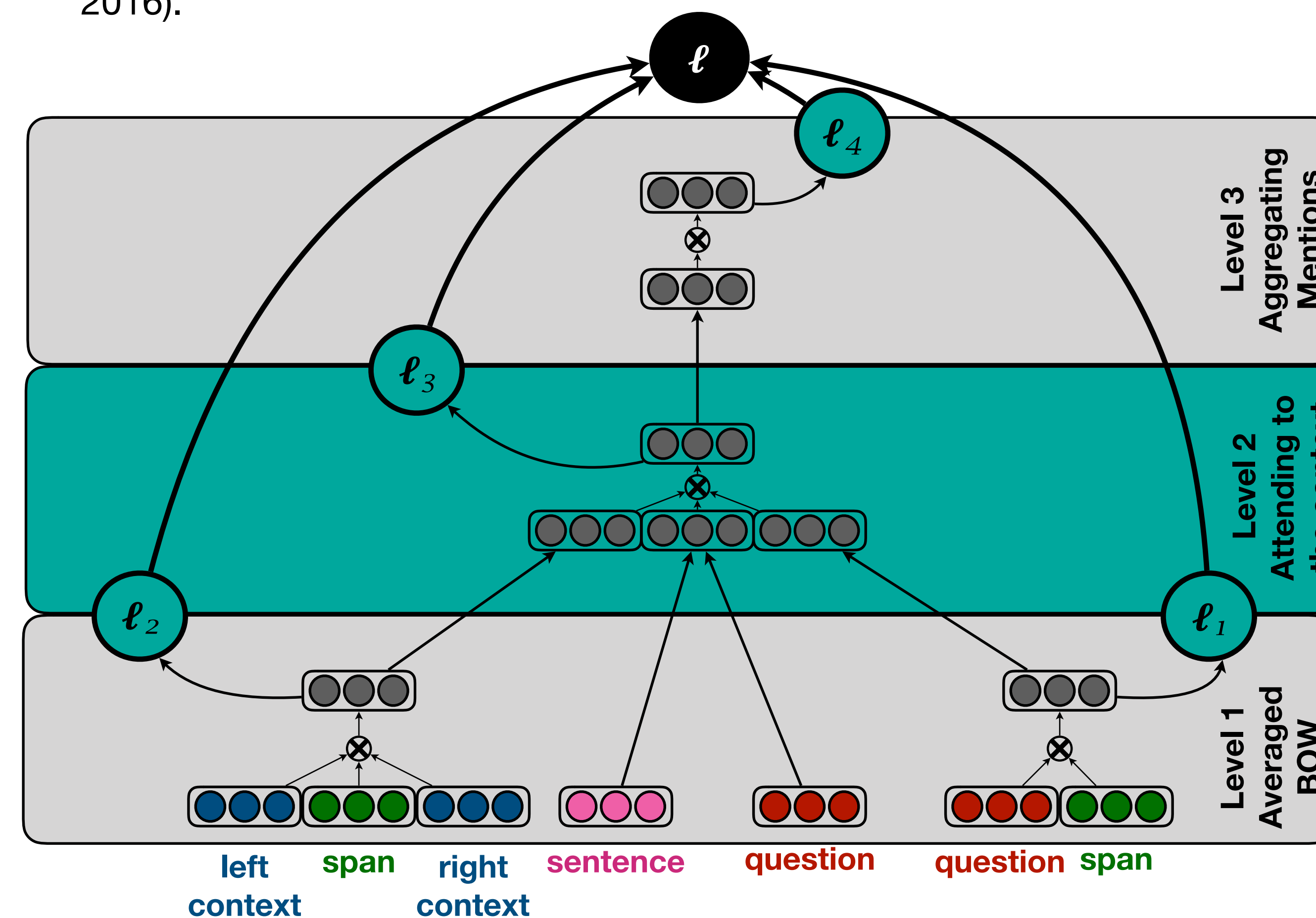- Need to **aggregate information** across multiple mentions.

Multiloss Architecture

Cascade of models arranged by levels,
- Level 1 : Averaged Bag of Words
  - Span in context
  - Question + Span
- Level 2 : Sentence in Context
- Level 3 : Aggregating Identical Mentions

Handling distant supervision:
- The exact occurrence of the correct span is treated as a **latent variable**.
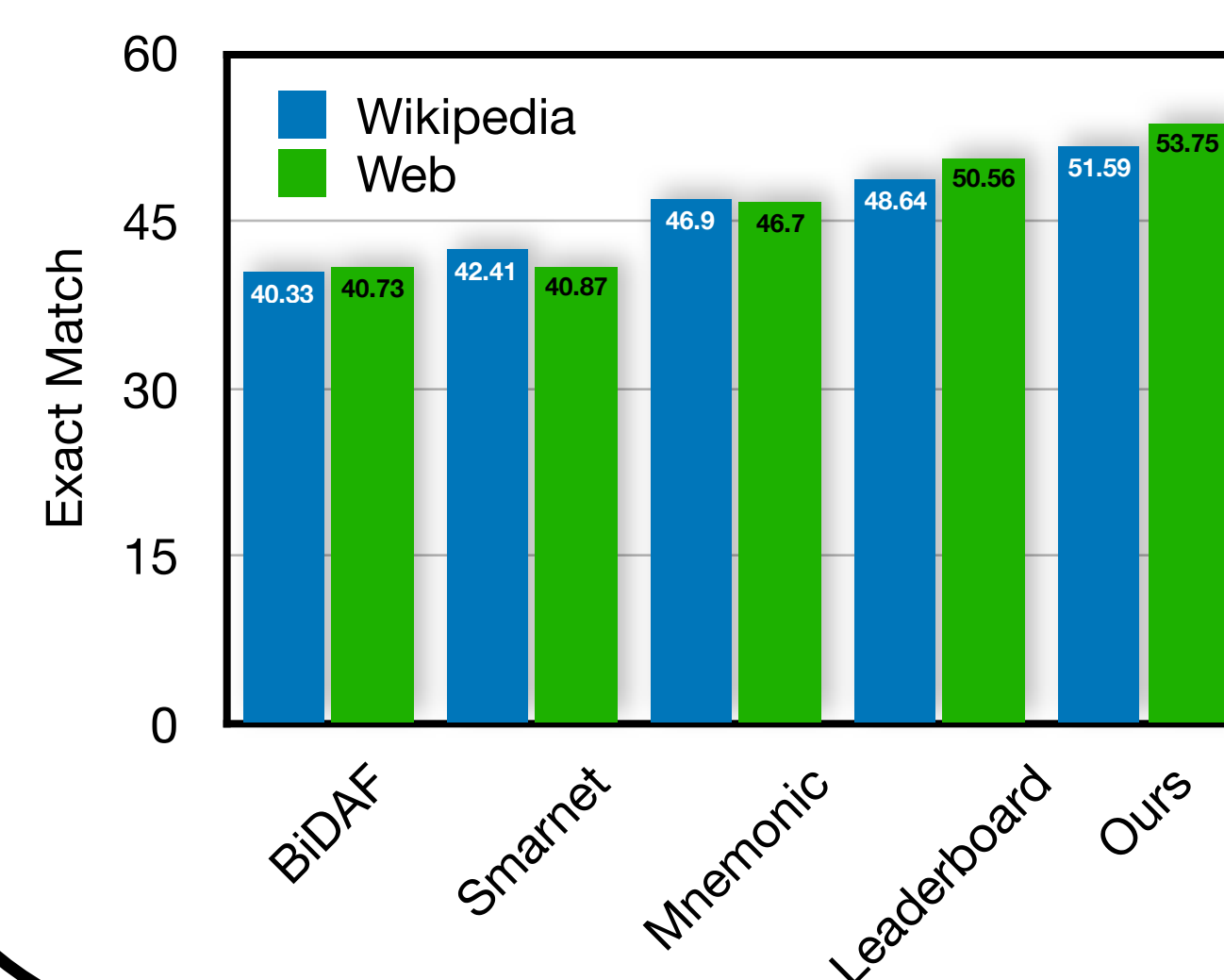- Probability of all occurrences maximized.

Multiloss Formulation: Each level is associated with its own loss (Al Rfou et. al., 2016).
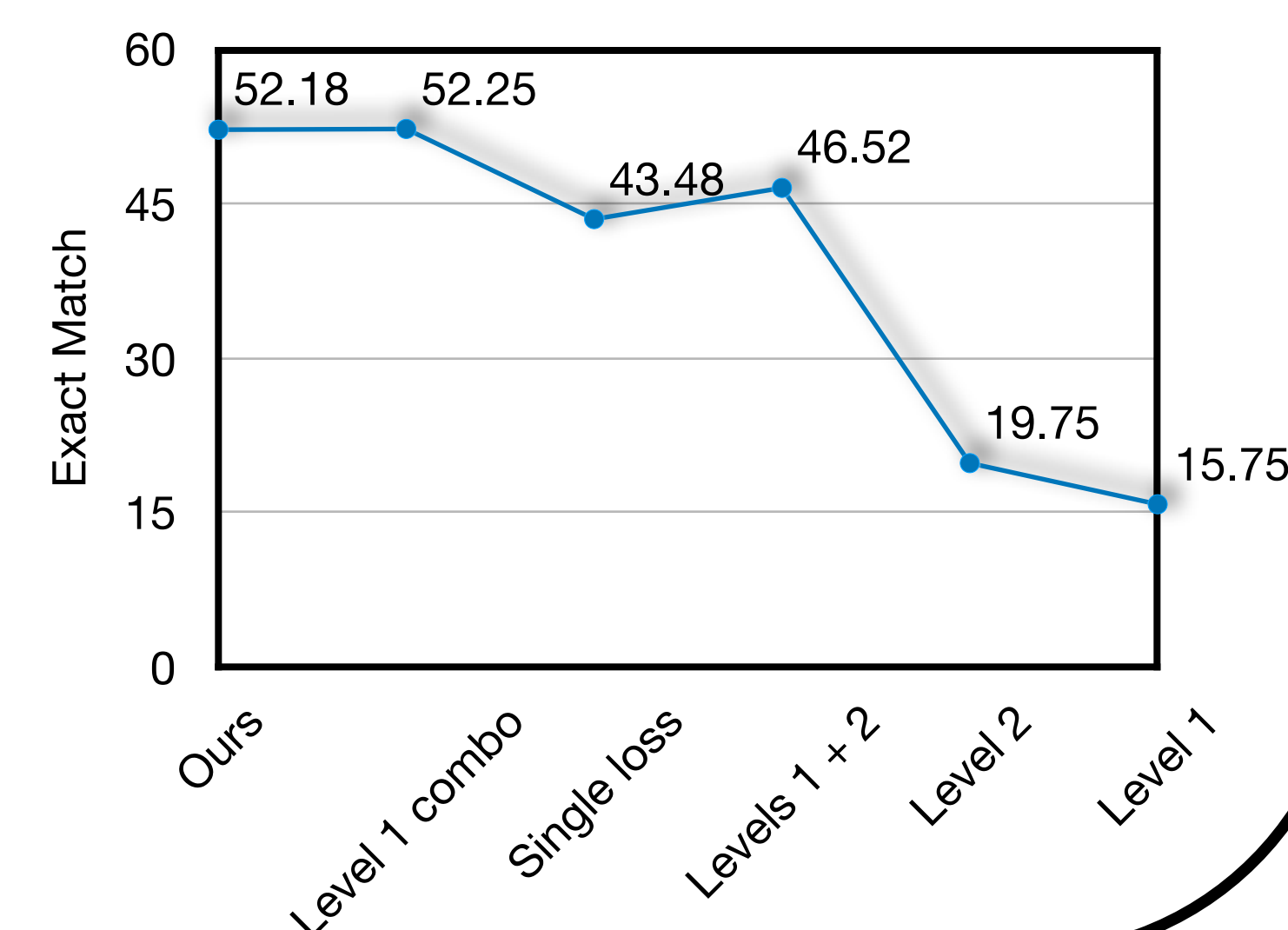
Level 3 Aggregating Mentions
Level 2 Attending to the context
Level 1 Averaged BOW

left context   span   right context   sentence   question   question   span

Richard Bachman is a pen name for fiction author Stephen King .   Which famous novelist … Bachman ?

Results on TriviaQA

TriviaQA Wikpedia and Web
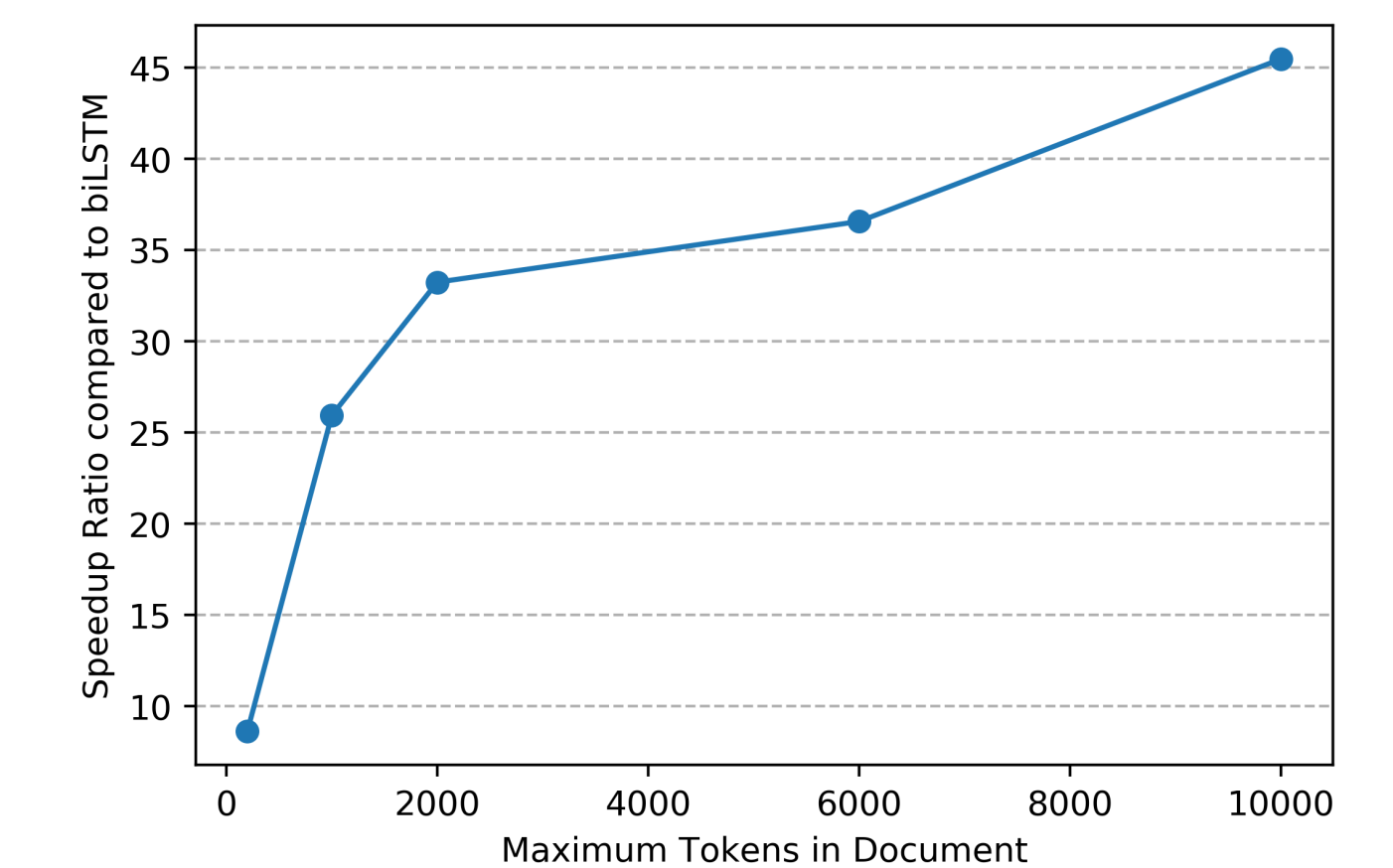
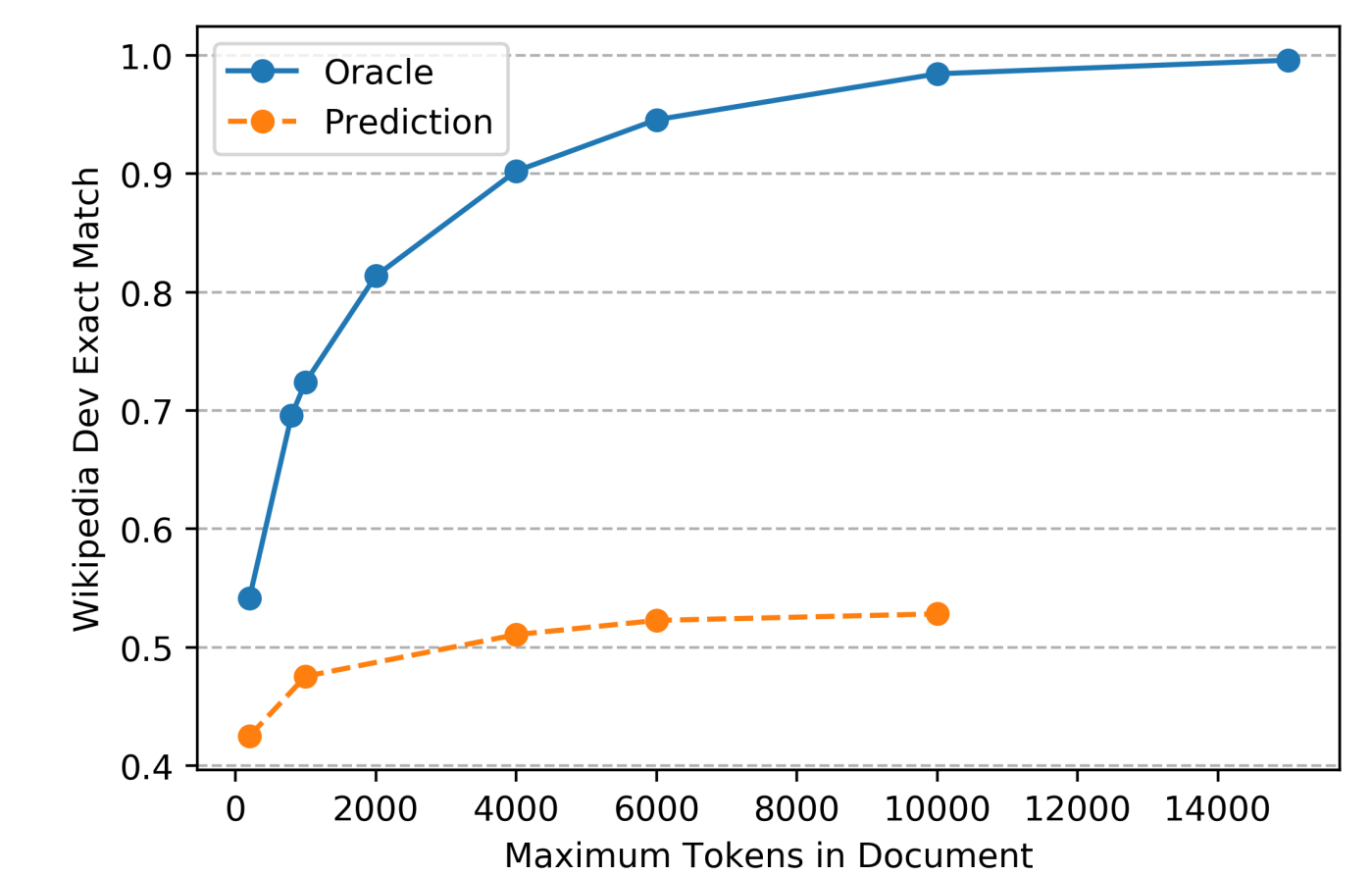|	Wikipedia	Web
BiDAF	40.33	40.73
Smarnet	42.41	40.87
Mnemonic	46.9	46.7
Leaderboard	48.64	50.56
Ours	51.59	53.75

Architecture Ablations

Ours 52.18, Level 1 combo 52.25, Single loss 43.48, Levels 1 + 2 46.52, Level 2 19.75, Level 1 15.75

Analysis of Performance

Speed vs LSTMs

Effect of Truncation

Oracle
Prediction

Top *k* performance

Level 3 (Ours)
Level 1 + Level 2 Only
Level 1 (Span + Context)
Level 1 (Question + Spans)
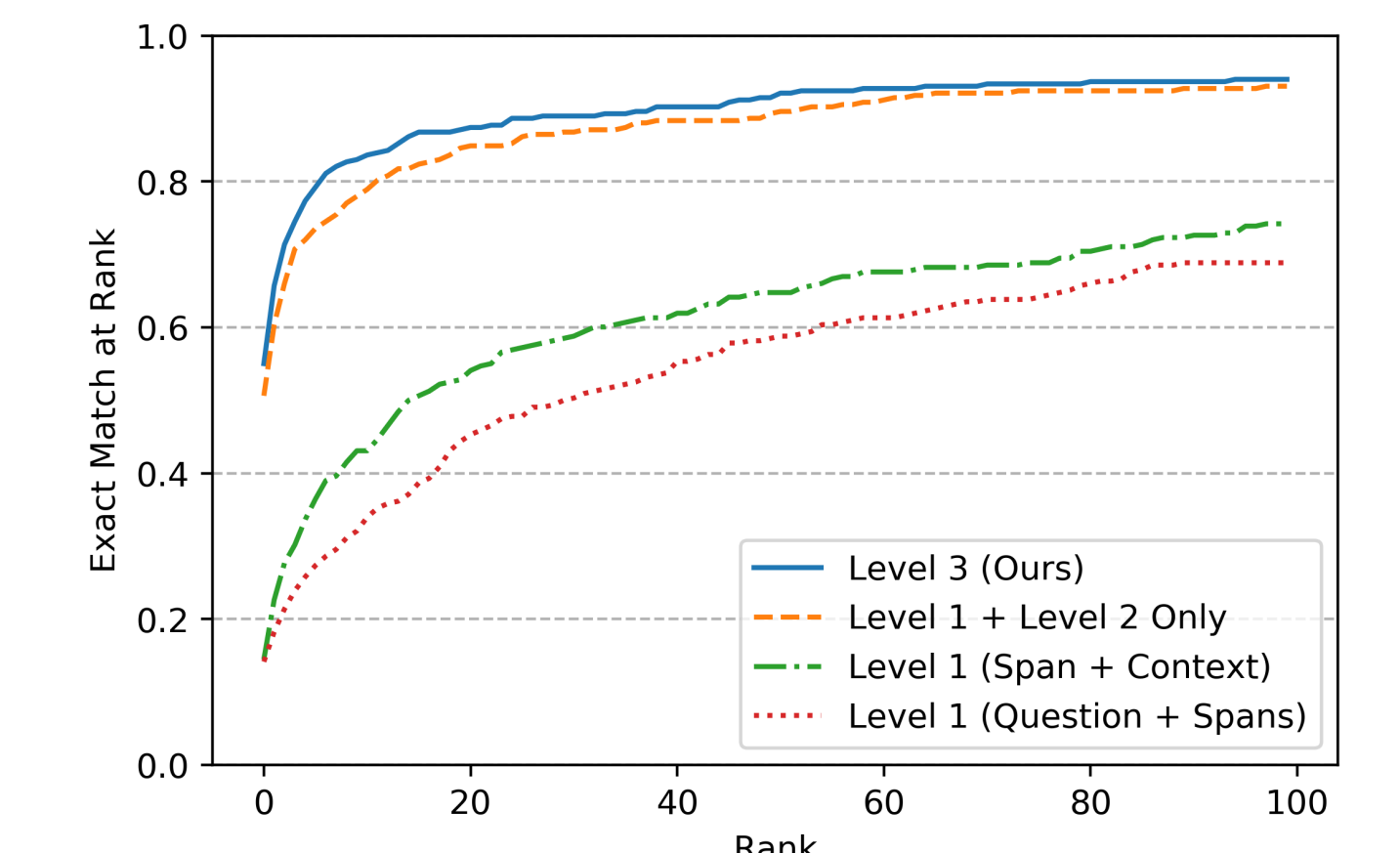
Take-home lessons

✓ Multiloss formulation helps.
✓ Aggregation is necessary.
✓ Context is important for comprehension.

Common Mistakes

✗ Confusing similar entities within a sentence.
✗ Question entities as answers.