

Introduction to NLP

CSE5321/CSEG321

Lecture 20. Retrieval-augmented Generation

Hwaran Lee (hwaranlee@sogang.ac.kr)

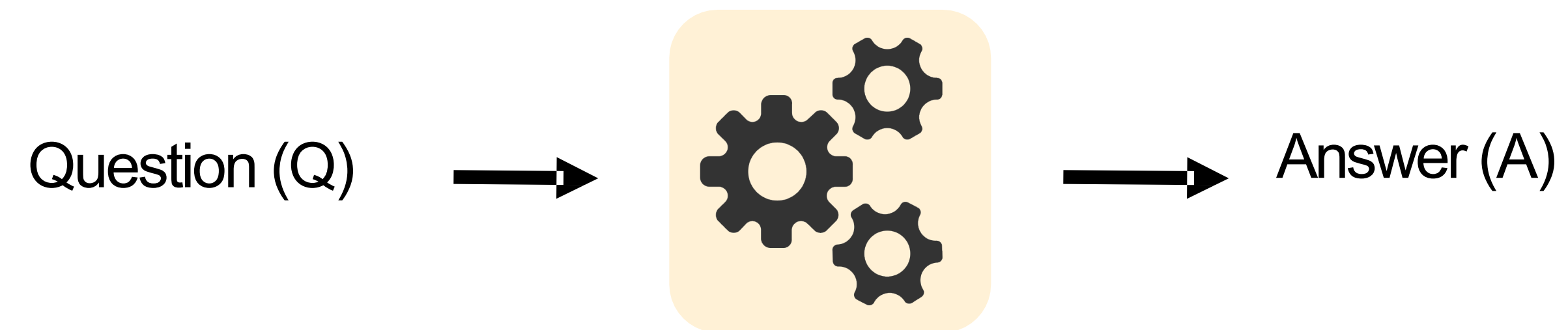
Lecture Plan

1. Question answering
2. Classes of QA: Closed-book, Open-book QA
3. Retrieval-augmented Generation (RAG)

Notice:

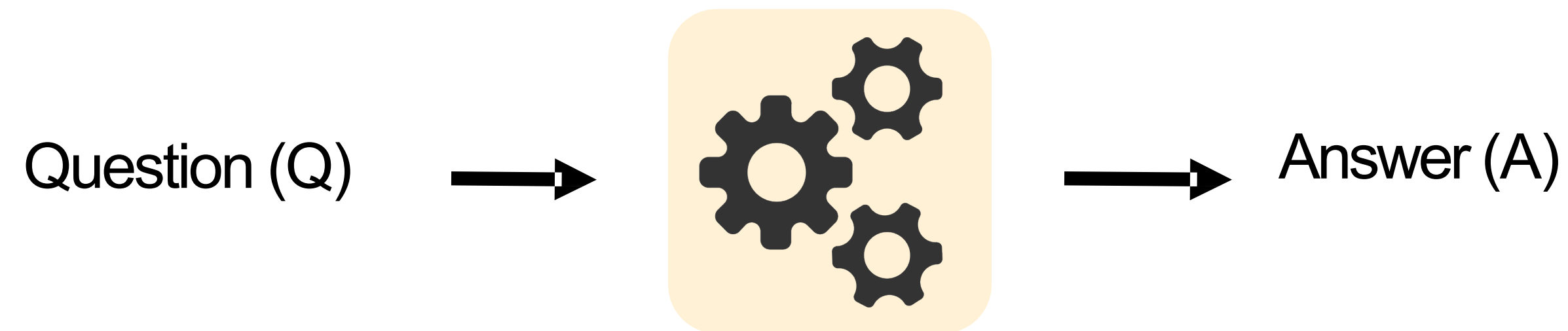
- Next lecture will be replaced with Colloquium Seminar
 - Date and Time: 5/28 AS 510
 - Title: Evaluation with socio-cultural awareness and multilingual LLMs

What is question answering?



The goal of question answering is to build systems that **automatically** answer questions posed by humans in a **natural language**

Question answering: a taxonomy



- **What information source does a system build on?**
 - A text passage, all Web documents, knowledge bases, tables, images..
- **Question type**
 - Factoid vs non-factoid, open-domain vs closed-domain, simple vs compositional, ..
- **Answer type**
 - A short segment of text, a paragraph, a list, yes/no, ...

Different types of QA

1. Closed book QA
 - Parametric QA
2. Open book QA
 - Reading Comprehension
 - Retrieval-augmented Generation (RAG)

Closed book QA

Closed book – answering the question with *no access to external information*

Q&A done just by ‘predicting the next word’

Seoul is the capital of _____

When does this work? Why do we care about this setting?

