

When Watson receives a question, the first step is question analysis.

One of the things Watson tries to figure out in this step is what the question is looking for.

This is defined as trying to find the Lexical Answer Type, or LAT, of the question.

The LAT is the word or noun in the question that specifies the type of answer.

You should be able to replace the LAT with the answer to complete the sentence.

For example, for the question, "Mozart's last and perhaps most powerful symphony shares its name with this planet," the LAT in this case is "this planet." If we replace this with the answer "Jupiter," it makes sense.

Mozart's last and perhaps most powerful symphony shares its name with Jupiter.

For the question, "Smaller than only Greenland, it's the world's second largest island," the LAT is "it's." If we replace the LAT with the answer "New Guinea," it makes sense.

"Smaller than only Greenland, New Guinea is the world's second largest island." Unfortunately, the LAT is not "island," which would be more descriptive, since the sentence with "New Guinea" in place of "island" does not make sense.

We can see in these two examples that sometimes the LAT is very specific, like "this planet," and sometimes it's very vague, like "it's." If we know the LAT, we know what to look for.

However, in an analysis of 20,000 questions, 2,500 distinct LATs were found, and 12% of the questions did not even have an explicit LAT.

They had LATs like "it's." Furthermore, even the most frequent 200 explicit LATs cover less than 50% of the questions.

So to enhance the question analysis step, Watson also performs relation detection to find relationships among words and decomposition to split the question into different clues.

The second step in Watson is hypothesis generation.

The goal of this step is to use the question analysis of step one to produce candidate answers by searching the databases.

In this step several hundred candidate answers are generated.

For the question, "Mozart's last and perhaps most powerful symphony shares its name with this planet," candidate answers could be Mercury, Earth, and Jupiter.

These are generated using various search techniques.

Then each candidate answer plugged back into the question in place of the LAT is considered a hypothesis.

For the question about Mozart's symphony, hypothesis one would be the question with "Mercury" in place of "this planet." Hypothesis two would have "Jupiter" in place of "this planet." And hypothesis three would have "Earth" in place of "this planet." If the correct answer is not generated at this stage, Watson has no hope of getting the question right.

Therefore, this step errors on the side of generating a lot of hypotheses and leaves it up to the next step to find the correct answer.

In the next video, we'll discuss how steps three and four score and rank the hypotheses.