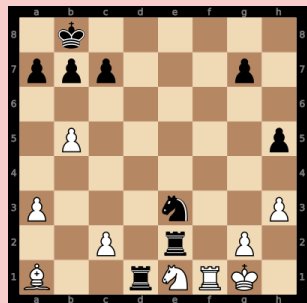


Chess



Text:

1k6/ppp3p1/8/1P5p/
8/P3n2P/2P1r1P1/
B2rNRK1 w - - 5 32

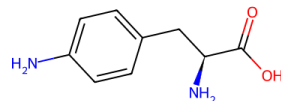
Fork Detection: Which of the following pieces is forking other pieces in this chess position?

Centipawn Eval: Which of the following is the correct centipawn evaluation for this chess position?

Puzzle Solving: Which of the following moves is the best next move for the active player in this chess position?

Legal Move Identification: Which of the following moves is legal in this chess position?

Chemistry



Text:

C1=CC(=CC=C1C
[C@@H](C(=O)O)N)
N.O

Carbon Counting: Which of the following is the correct number of carbon atoms in this compound

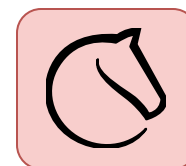
Hydrogen Counting: Which of the following is the correct number of hydrogen atoms in this compound?

Weight Calculation: Which of the following is the correct molecular weight of this compound?

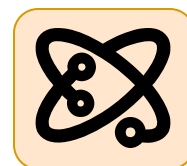
Caption Matching: Which of the following descriptions is correct for this compound?

Benchmark Construction

1. Sampling data from existing dataset/tools



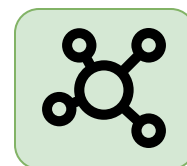
Lichess



ChemQA



Irishman



NetworkX

2. Sampling Raw Text Representation

FEN

3r4/rR1k1...

SMILES

C1=CC(=C-...

ABC

X:1829nL:...

Adj Matrix

[[0, 1, 0, 0, 1], ...

3. Determining Ground Truth Answers

Best Move: d8d3
Fork: White's Queen on d8
...

Carbon: 9
Weight: 180
...

Form: American Popular
Note Count: 8
...

Path 1 -> 4: 3
Shortest Path 3 -> 4: 2
...

4. Crafting Incorrect Choices

Numerical Choice

Offset correct answer for plausible, challenging options.

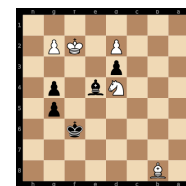
Categorical Choice

Sample incorrect options uniformly from candidates.

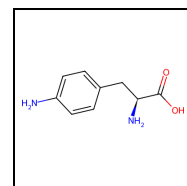
Descriptive Choice

Sample semantically similar alternatives using E5 model.

5. Converting to Standard Image Representations



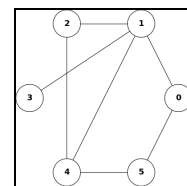
python-
chess



rdkit



music21



NetworkX

Music



Text:

X:1829nL:1/16
nM:2/4nK:G\n
(3(BcB) A2d>c ... I]

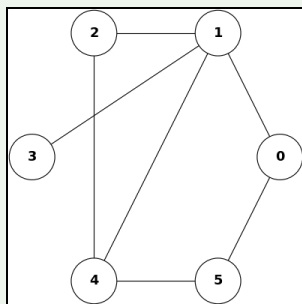
Notes Counting: How many individual *note* notes (including #, b, and) appear in this piece?

Measure Counting: Which of the following is the correct number of measures in this piece?

Forms Identification: Which of the following best describes the musical form of this piece?

Rhythm Identification: Which of the following measures contains a *rhyme pattern*?

Graph



Text:

[[0, 1, 0, 0, 0, 1],
[1, 0, 1, 1, 1, 0],
[0, 1, 0, 0, 1, 0],
[0, 1, 0, 0, 0, 0],
[0, 1, 1, 0, 0, 1],
[1, 0, 0, 0, 1, 0]]

Path Counting: Which of the following is the correct number of unique simple paths from a to b in the graph?

Path Matching: Which of the following node lists represents a path from a to b?

Shortest Path Identification: Which of the following is the length of the shortest simple path from a to b?

BFS Traversal: Which of the following lists represents the order of the BFS traversa from node a in the graph?