Definition: An entity is an ORGANISM (all individual life forms such as microorganisms, plants, animals, mammals, insects, fungi, bacteria etc.), PHENOMENA (occurring natural, biological, physical or chemical processes such as decomposition, colonisation, deforestation, as well as events, such as climate change etc.), MATTER (chemical and biological compounds, and natural elements, such as carbon, sediment, sand etc.), ENVIRONMENT (natural and man-made environments organisms live in, such as groundwater, garden, aquarium, mountain etc.), QUALITY (data parameters measured or observed, phenotypes and traits, such as volume, age, structure, morphology etc.), and LOCATION (geographic location such as China, the United States etc.). Dates, times, and adjectives are not entities.

Named entity categories, definitions, and real-world examples.

Example 1: [[0, 'Because'], [1, 'fungal'], [2, 'pathogens'], [3, 'likely'], [4, 'have'], [5, 'similar'], [6, 'abiotic'], [7, 'requirements'], [8, 'for'], [9, 'growth'], [10, 'as'], [11, 'other'], [12, 'fungi'], [13, ','], [14, 'characterizing'], [15, 'weather'], [16, 'conditions'], [17, 'favorable'], [18, 'for'], [19, 'fungi'], [20, 'also'], [21, 'may'], [22, 'be'], [23, 'used'], [24, 'to'], [25, 'predict'], [26, 'the'], [27, 'selective'], [28, 'pressures'], [29, 'imposed'], [30, 'by'], [31, 'pathogenic'], [32, 'fungi'], [33, 'on'], [34, 'plants'], [35, 'in'], [36, 'different'], [37, 'habitats'], [38, '.']]

Answer: [['growth', 'PHENOMENA', 9, 9], ['fungi', 'ORGANISM', 12, 12], ['weather conditions', 'QUALITY', 15, 16], ['fungi', 'ORGANISM', 19, 19], ['fungi', 'ORGANISM', 32, 32], ['plants', 'ORGANISM', 34, 34], ['habitats', 'ENVIRONMENT', 37, 37]]

Example 2: [[0, '--'], [1, '-'], [2, '6'], [3, 'digit'], [4, 'metal'], [5, 'tags'], [6, 'starting'], [7, 'with'], [8, '3'], [9, 'were'], [10, 'also'], [11, 'used'], [12, 'for'], [13, 'woody'], [14, 'debris'], [15, 'items'], [16, 'CSP'], [17, 'metal'], [18, 'tag'], [19, 'number'], [20, '('], [21, 'trees'], [22, ','], [23, 'woody'], [24, 'debris'], [25, ')'], [26, '('], [27, 'TagMBa'], [28, ')'], [29, ','], [30, 'dimensionless'], [31, 'TagMBa'], [32, 'CSP'], [33, 'tree'], [34, 'individuals'], [35, 'were'], [36, 'marked'], [37, 'mostly'], [38, 'with'], [39, 'metal'], [40, 'tags'], [41, 'but'], [42, 'also'], [43, 'additional'], [44, 'tags'], [45, 'were'], [46, 'used'], [47, '.']]

Answer: [['metal', 'MATTER', 4, 4], ['woody', 'ENVIRONMENT', 13, 13], ['metal', 'MATTER', 17, 17], ['trees', 'ORGANISM', 21, 21], ['woody', 'ENVIRONMENT', 23, 23], ['tree', 'ORGANISM', 33, 33], ['metal', 'MATTER', 39, 39]]

Example 3: [[0, '('], [1, 'Phenolics'], [2, ':'], [3, 'Total'], [4, 'phenolics'], [5, 'content'], [6, 'as'], [7, 'tannic'], [8, 'acid'], [9, 'equivalent'], [10, ')'], [11, 'dimensionless'], [12, 'real'], [13, 'Secondary'], [14, 'Metabolites'], [15, 'Secondary'], [16, 'metabolites'], [17, 'are'], [18, 'organic'], [19, 'compounds'], [20, 'that'], [21, 'are'], [22, 'not'], [23, 'directly'], [24, 'involved'], [25, 'in'], [26, 'the'], [27, 'normal'], [28, 'growth'], [29, ','], [30, 'development'], [31, ','], [32, 'or'], [33, 'reproduction'], [34, 'of'], [35, 'an'], [36, 'organism'], [37, '.']]

Answer: [['tannic acid', 'MATTER', 7, 8], ['Secondary Metabolites', 'MATTER', 13, 14], ['Secondary metabolites', 'MATTER', 15, 16], ['organic compounds', 'MATTER', 18, 19], ['normal growth', 'PHENOMENA', 27, 28], ['development', 'PHENOMENA', 30, 30], ['reproduction', 'PHENOMENA', 33, 33], ['organism', 'ORGANISM', 36, 36]]

Examples of the task. The LM is given a list of indices and tokens, and is expected to retrieve the NE instance, NE category, and NE indices.

In this example, the model is presented with 3 random example sentences.

Generate ONLY a Python list with a nested list of named entities from the sentence: [[0, 'The'], [1, 'primacy'], [2, 'of'], [3, 'either'], [4, 'species'], [5, 'or'], [6, 'functional'], [7, 'group'], [8, 'richness'], [9, 'effects'], [10, 'depended'], [11, 'on'], [12, 'the'], [13, 'sequence'], [14, 'of], [15, 'testing'], [16, 'these'], [17, 'terms'], [18, ','], [19, 'indicating'], [20, 'that'], [21, 'both'], [22, 'aspects'], [23, 'of], [24, 'richness'], [25, 'were'], [26, 'congruent'], [27, 'and'], [28, 'complementary'], [29, 'to'], [30, 'expected'], [31, 'strong'], [32, 'effects'], [33, 'of], [34, 'legume'], [35, 'presence'], [36, 'and'], [37, 'grass'], [38, 'presence'], [39, 'on'], [40, 'plant'], [41, 'chemical'], [42, 'composition'], [43, '.']]. one-token entity: [entity, label, index of token in list, index of token of list] multi-token entity: [entity, label, index of first token, index of last token] DO NOT HALLUCINATE

Instructions: The LM needs to follow the output guidelines and generate a nested Python list.