**プログラム処理フロー（概要）**

1. **クエリと論文の類似度取得**
   * **summary列に基づいて全論文の要約をリスト化し、クエリ文字列を追加して埋め込みを取得。**
   * **cosine\_similarityを用いてクエリと各論文の要約埋め込みとのコサイン類似度を計算。**
   * **類似度スコアをsimilarity\_query列に追加し、正規化したスコアをquery\_similarity列として保存。**
2. **ノードごとの上位類似度作成・時系列順に並び替え**
   * **published\_dayを基に発行日順でソート。**
   * **time\_series\_similarity関数で発行日以降の上位5件の類似論文を取得し、スコアとタイトルを新規列に追加。**
3. **similar\_node\_dicの作成**
   * **上位類似度データを基に類似度が閾値を超えるペアを判定し、親子関係（similar\_parentとsimilar\_child）として保存。**
   * **発行日が将来の論文を対象とするエッジは除外。**
4. **グラフの生成およびプロット**
   * **時系列グラフ（similar\_node\_dic使用）: 閾値を超える類似論文ペアにエッジを追加し、最も関連の深い論文の類似関係を時系列順に表示。**
   * **階層構造グラフ（node\_dic使用）: クエリからの深さに応じた階層構造をエッジ色で表現し、発行年順の親子関係を強調。**

**プログラム処理フロー（詳細）**

1. **クエリと論文の類似度取得**
   * **処理内容:**
     + **summary列に基づいて全論文の要約をリスト化し、これにクエリ文字列を追加して埋め込みを取得。**
     + **cosine\_similarityを用い、クエリと各論文の要約埋め込みとのコサイン類似度を計算。**
     + **クエリとのコサイン類似度スコアをsimilarity\_query列に追加し、正規化した類似度スコアをquery\_similarity列として保存。**
   * **出力データ: similarity\_queryとquery\_similarity列が追加されたDataFrame（クエリと各論文の類似度データ）**
2. **ノードごとの上位類似度作成・時系列順に並び替え**
   * **処理内容:**
     + **published\_dayを年単位でソート。**
     + **time\_series\_similarity関数で、各論文に対して発行日以降で最も類似する5つの論文を取得。**
     + **上位5つの類似スコア（similarity\_1〜similarity\_5）とタイトル（title\_similarity\_1〜title\_similarity\_5）として新規列に保存。**
   * **出力データ: 各論文に対し、上位5つの類似論文とそのスコアを含むDataFrame**
3. **similar\_node\_dicの作成**
   * **処理内容:**
     + **上位類似度データを基に、各論文に対して類似度が閾値（similarity\_threshold）を超えるかを判定。**
     + **閾値を超える類似論文ペアについて、親子関係をsimilar\_parentとsimilar\_childとしてsimilar\_node\_dicに格納。**
     + **発行日が将来の論文を対象としたエッジは除外。**
   * **出力データ: similar\_node\_dic DataFrameは、閾値を満たす親子論文ペアを格納。**
4. **グラフの生成およびプロット**
   * **処理内容:**
     + **ノードの位置決定: 各ノード（論文）をカテゴリ別、年代別、著者別のプロット位置に配置。**
     + **エッジの設定:**
       - **\*\*時系列グラフ（similar\_node\_dic使用）\*\*では、similar\_node\_dicの親子関係に基づき、類似度が閾値を超える論文間にエッジを追加し、類似関係を表示。**
       - **\*\*階層構造グラフ（node\_dic使用）\*\*では、発行日順での階層構造に基づき、特定の深さに基づいたエッジを色分けして表示。**
     + **ノードの表示方法:**
       - **カテゴリ別表示: 論文カテゴリごとに異なる色でノードを表示。コンピュータサイエンス（赤）、数学（オレンジ）、バイオサイエンス（茶色）など。**
       - **年代別表示: 発行年に基づき特定の年代のノードを強調して表示。**
       - **著者別表示: 論文の著者ごとにノードをフィルタリングし、著者間の関係を確認可能。**
     + **エッジのインタラクティブ表示:**
       - **各エッジは論文間の類似度や親子関係を反映し、インタラクティブメニューでカテゴリや年代、著者ごとのエッジ表示を切り替え可能。**

**入出力データと中間データのフォーマット（詳細）**

* **クエリと論文の類似度計算結果**
  + **形式: DataFrame**
  + **列:**
    - **title: 論文のタイトル**
    - **similarity\_query: クエリとの類似度**
    - **query\_similarity: 正規化された類似度スコア**
* **時系列上位類似度データ**
  + **形式: DataFrame**
  + **列:**
    - **similarity\_1〜similarity\_5: 上位5つの類似度スコア**
    - **title\_similarity\_1〜title\_similarity\_5: 上位5つの類似論文タイトル**
* **similar\_node\_dic**
  + **形式: DataFrame**
  + **列:**
    - **similar\_parent: 親ノード（類似性の基準となるタイトル）**
    - **similar\_child: 子ノード（親と類似度が閾値を超えたタイトル）**
* **node\_dic**
  + **形式: DataFrame**
  + **列:**
    - **parent: 親ノード（基準となる論文タイトル）**
    - **child: 子ノード（親と関連するタイトル）**
    - **depth: 階層の深さ（クエリからの距離）**

**二つのグラフの違い（詳細）**

1. **時系列グラフ（similar\_node\_dic使用）**
   * **使用データ: similar\_node\_dic**
   * **目的: 論文の類似関係を可視化し、閾値に基づく類似論文ペアを表示。**
   * **エッジの設定:**
     + **類似度が閾値を超えるノードペア間にのみエッジを追加し、発行日が過去から未来へと続く方向で表示。**
     + **閾値を超えた論文間のみを結ぶことで、時系列に基づく最も関連の深い論文ペアが視覚的に際立つ。**
   * **表示モード:**
     + **カテゴリ、年代、著者ごとにフィルタリング可能で、発行年に基づく階層構造が強調される。**
     + **特定の分野や著者ごとの類似関係が詳細に表示され、最も関連性のある論文ペアのつながりを確認可能。**
2. **階層構造グラフ（node\_dic使用）**
   * **使用データ: node\_dic**
   * **目的: クエリから派生する論文の階層的親子関係を表示し、深さ情報に基づく階層構造を可視化。**
   * **エッジの設定:**
     + **depth列を参照し、クエリからの階層ごとに異なる色のエッジを追加。**
     + **各エッジは発行日順での階層関係に従い、クエリからの距離に応じてグラフ上に配置される。**
   * **表示モード:**
     + **カテゴリ、年代、著者ごとのフィルタリングが可能で、年代を越えた階層的関係を視覚化。**
     + **特定分野内でのクエリからの階層構造に基づく関係性を詳細に表示できる。**