

RAG

知識基盤・指令塔

底生動物と底質環境間・相互作用のモデル化
知識DBとRAGによる環境データ解析支援

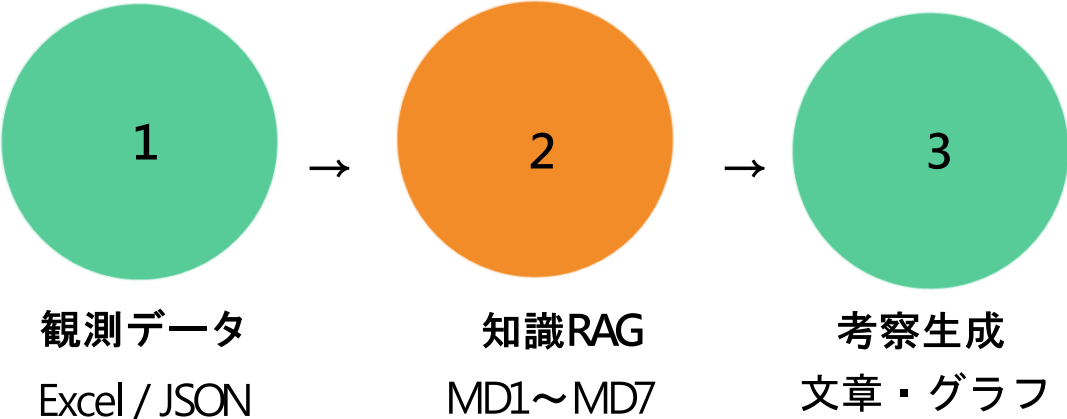
広島大学 日比野忠史

AI

実行エンジン

生成AI時代を見据えた海象学習モデルの構築

文書検索ではなく、考察知識をデータに適用するRAG



図の読み方を教えるRAG

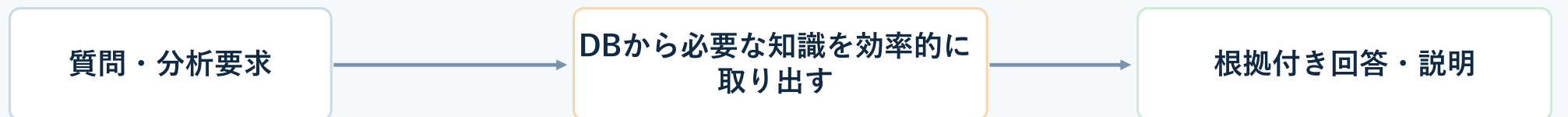
MD（マークダウン）はコンピュータにも人にもわかるファイル形式を表す拡張子名（.md）

表 作成した考察知識のRAG（MD1~MD7）一覧

id	MD名	内容
1	C/IL比による有機物分解状態	C/IL比を使って、未分解BOM（約0.4）・一般底泥（約0.2）・難分解化（0.2未満）を判定する基準.
2	M/A比（軟体/節足）による優占構造	M/A比で軟体動物優占か節足動物優占かを読む。低値域（0.001など）の扱いも定義.
3	TOCとlog(M/A)に基づく競合勾配	TOC増加に対する群集変化をlog(M/A)で解釈。低TOC域の直線関係や高TOC域の制限傾向を扱う.
4	TOC(砂)によるOSS量評価	TOC(砂)を、細粒分に吸着しない砂場由来有機物（OSS）の量指標として解釈する.
5	細粒分・D50と有機物蓄積	細粒化（高細粒分・低D50）とTOC増加、生息制限との関係を読むための知識.
6	土粒子密度から読む有機物混入・難分解化の解釈指針	土粒子密度の変化を、有機物混入状態や分解進行（鉱物結合・高密度化）と結びつけて解釈.
7	D50・IL・土粒子密度の統合解釈	D50・IL・土粒子密度を統合し、砂場/泥場やOSS混入などを4分類で判定する総合ルール.

RAGとは何か：DB・RAG・LLMは別物

一般的なRAGの基本構造



RAGなし
LLMの内部知識・推測に依存しやすい

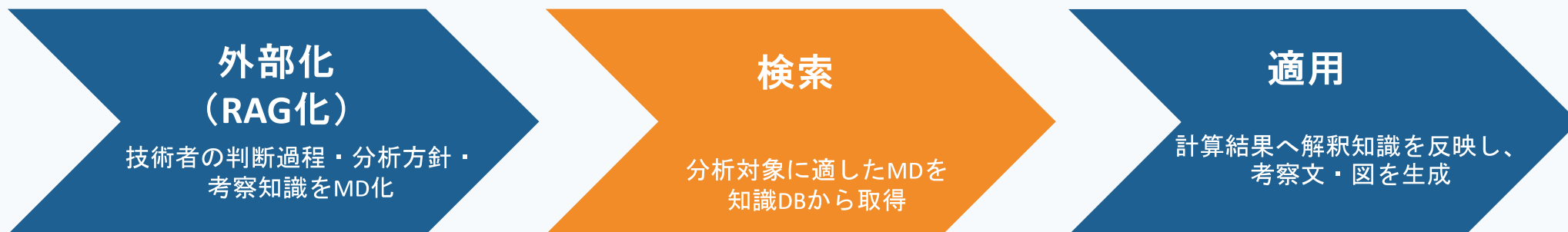
RAGあり
外部DBの知識を根拠として使える

要点：MD、Excel等を格納する保管庫、RAGは構造化されたDBから知識を探して（ルーティング）LLMに渡す仕組み、LLMは生成するAI本体

本研究でのRAGの位置付け

考察知識の適用エンジン

観測データから結論を直接予測するのではなく、
技術者が技術者の判断知識をMD化し、RAGで検索可能にした知識を検索し、
数値分析結果に適用する



従来型RAG：文書を探して答える

本研究：判断基準を探してデータに適用する

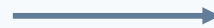
MD（マークダウン）はコンピュータにも人にもわかるファイル形式を表す拡張子名（.md）

この研究でのRAGは従来型と何が違うか

一般的なRAG

質問
↓
文書DB検索
↓
回答生成

文書検索付きチャット



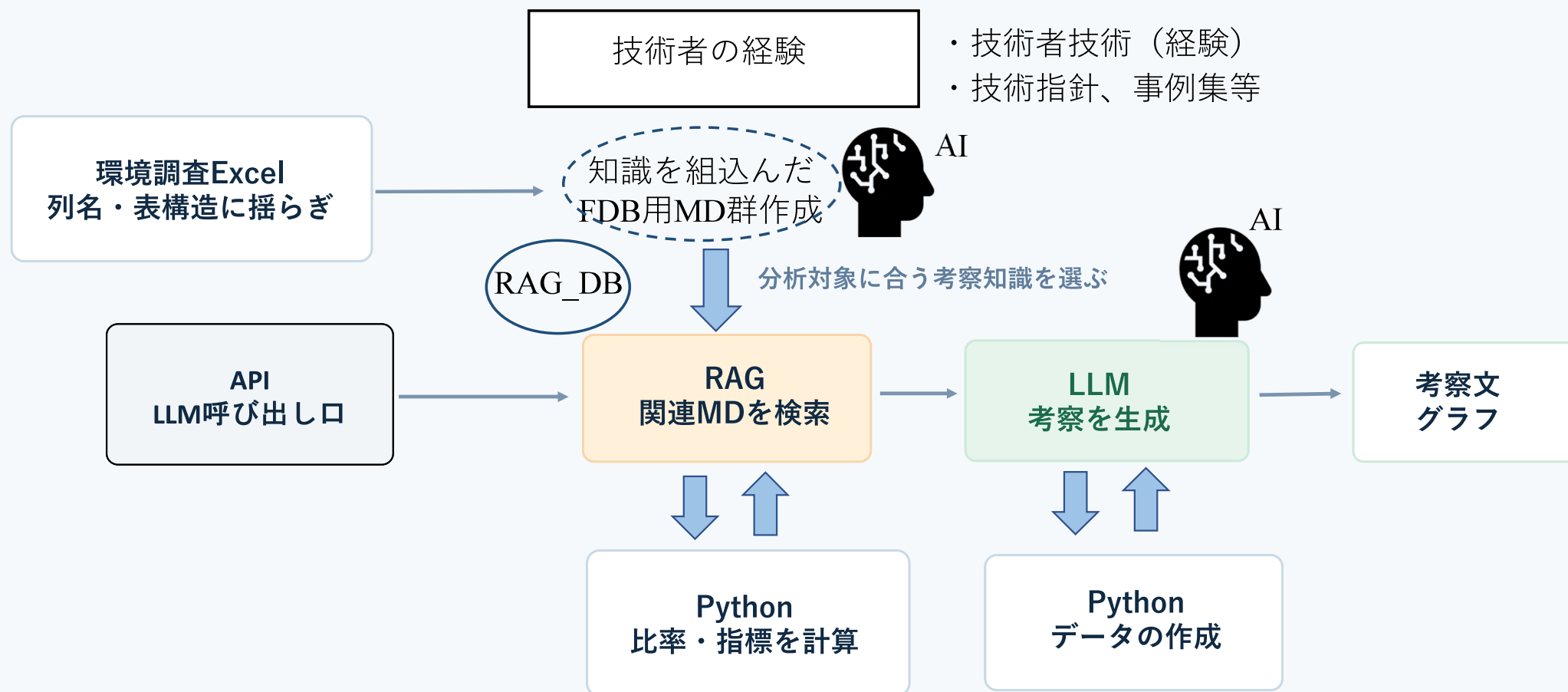
このRAG

Excel調査データ
↓
AIによる列名読み替え・JSON化
↓
Python計算
↓
RAGによる考察知識MD検索
↓
LLMによる考察生成・グラフ方針生成

考察知識適用型RAG

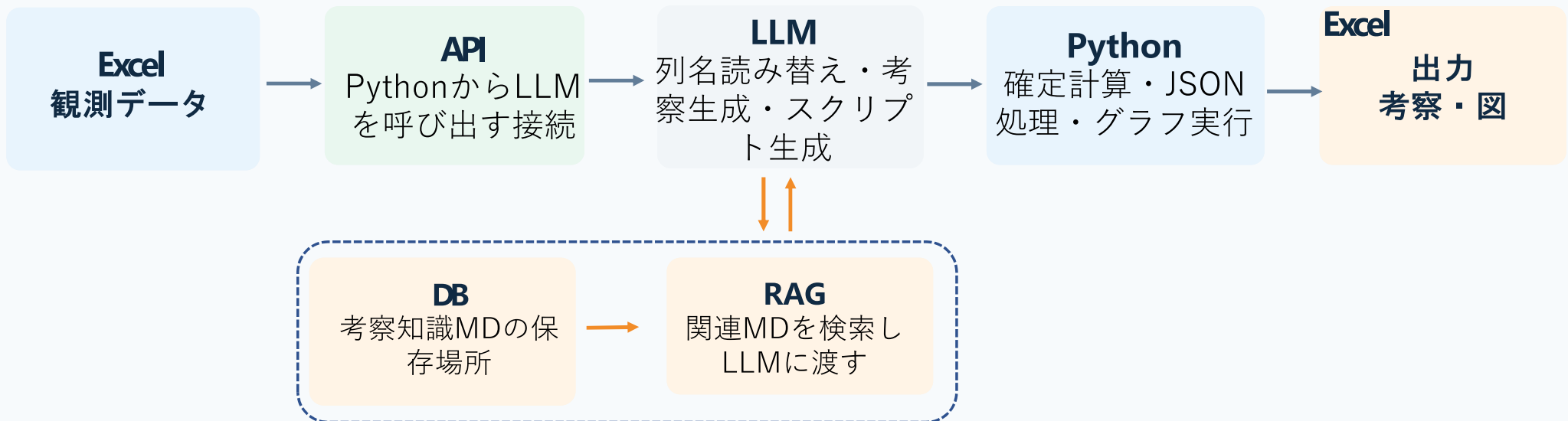
違いの本質：検索する対象が「文書」ではなく、観測データを読むための「判断基準」

この研究でのRAG：考察知識（人の知見）をルーティング（将来的にFDB化）



FDB（Federated DB）：複数のDB/RAGを横断して探せるしくみ

LLM・API・DB・RAG・FDB・Python・Excelの関係



DBは保存場所、**RAG**は検索・受け渡し、**LLM**は生成、**API**は接続口、**Python**は確定計算

FDBは複数DB/RAGを横断する発展形

DBとRAG（FDB）の関係

この研究では、DBに保存されるのは単なる文章ではなく、
底生動物・底質環境を解釈するための判断基準



DBだけでは答えを
生成しない

RAGはDBを検索して
LLMへ渡す手順

LLMが受け取った知識を
使って説明を作る

FDBは複数のDB/RAGを横断的につなぐ上位構造

なぜRAG（Retrieval - Augmented Generation / 検索-拡張生成）が必要か

底生動物
生息量・群集構造



底質環境
TOC・粒径・IL等

相互作用で関係が一義的に定まらない

観測データだけでは
複数解釈が残る

分析方法・考察が
分析者の経験に依存する

未経験者には
判断過程が見えにくい

必要なRAG：分析者の経験に依存していた前提・手続き・判断基準を外部化し、調査データ解析に再利用する仕組み

AIは元素比・強熱減量・燃焼温度（人の知識）を使うためにはRAGが必要

AIは元素比・強熱減量・燃焼温度を使える。ただし、AI単体に任せるのではなく、RAG・DBに「使うべき判断基準」として登録することが前提

AI単体にExcelだけを渡す場合

TOC・IL・粒径などを表面的に整理しやすい
C/IL比、燃焼温度、BOMの意味づけは不安定
「有機汚濁」「TOCが高い」程度で止まりやすい



RAG + DBに判断基準を入れる場合

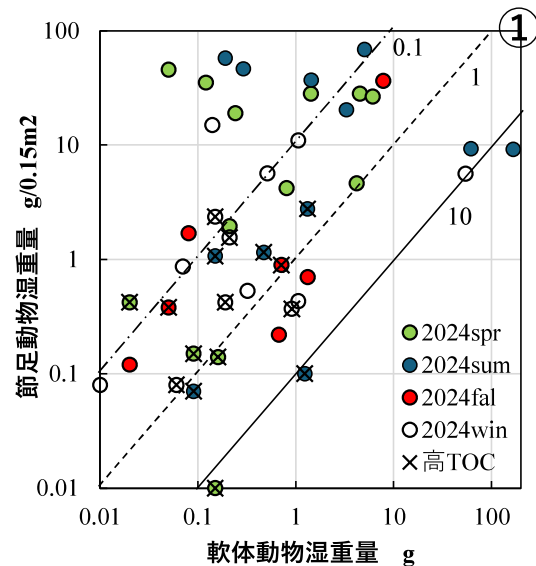
C/IL比、強熱減量、燃焼温度を確認する
D50・細粒分・土粒子密度と統合して読む
新生BOM・難分解化・OSSなどの候補を根拠付きで推定

この追加の意味

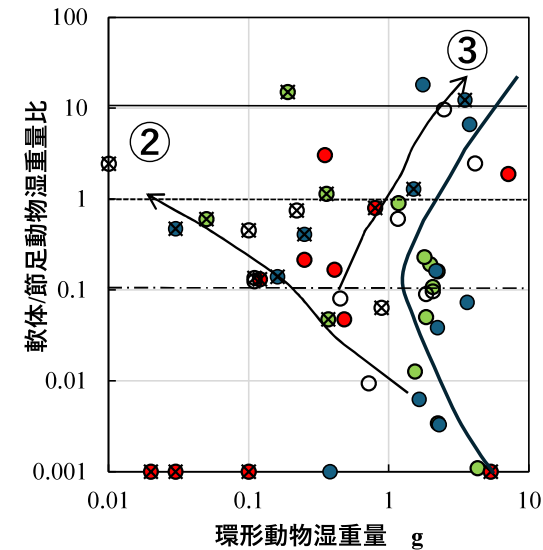
一般的な調査員やAI単体では見落としやすい化学的手掛かりを、AIが毎回参照する標準の読み方に変換できる

生物の特性を考えるRAGを作成する (MD1)

生物調査結果の分析は簡単ではない→技術者の知恵をRAG化する



M/A比は優占構造・競合関係の指標



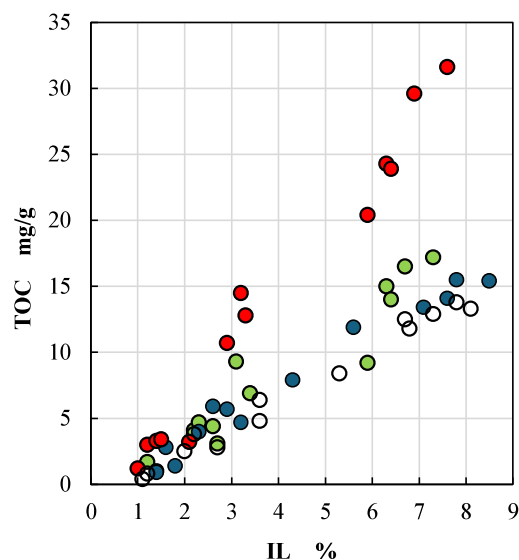
軟体動物湿重量と節足動物湿重量比較する意義、
軟体/節足動物湿重量比 (M/A比) を考える意義は何か？
軟体/節足動物湿重量比と環形動物湿重量を比較すると何がわかるか？

人が判断する

- ①軟体動物が増える場で節足動物が増える傾向
- ②節足動物の減少場で環形動物が増える
- ③軟体動物の増加場では環形動物の減少は無い

泥調査データの見方（はずれ値の意味）

知識MDの作成（MD6）

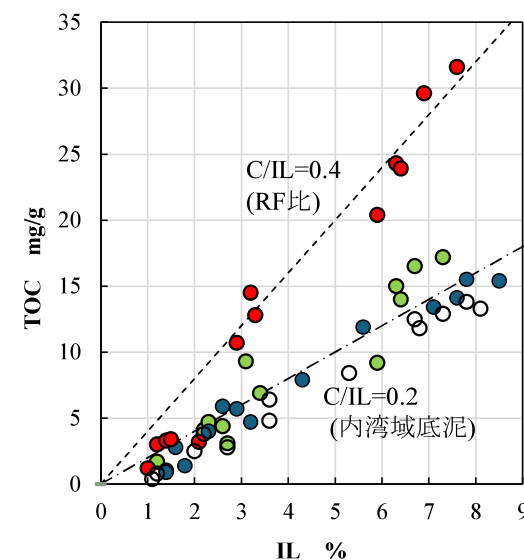


直線に乗らない
↓
外れ値の意味？

有機泥に関する知識の組み込み

- ①有機物はC, H, N, S, Oからなり，燃えると蒸発する
有機物が完全燃焼すると， H_2O , CO_2 , NO_2 , SO_2 が蒸発してP, Fe等の金属が残る
燃焼減量（IL）は蒸発元素C, H, N, S, Oの和

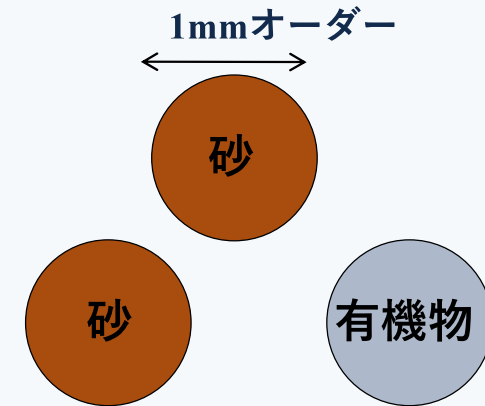
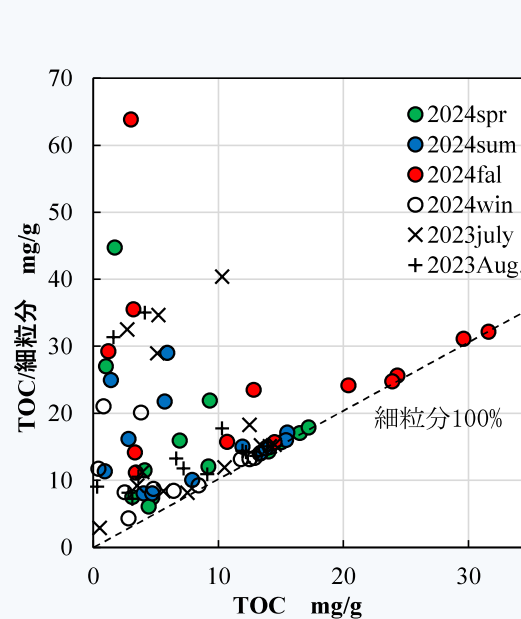
- ②レッドフィールド比
((CH_2O)₁₀₆(NH_3)₁₆(H_3PO_4)) を仮定すれば，生物により生成されたBOM（乾燥状態）のC/IL比
($IL = C + H + N + O + S$) は約0.4 (g-C/g-IL)



補助線が意味になる
↑
経験が必要

堆積砂泥の特性を考えるRAGを作成する (MD7)

有機物は細粒分に吸着し易いという知識（知らない人は知らない）を使う



砂粒子内では有機物は単独に存在

有機物は生物が生成

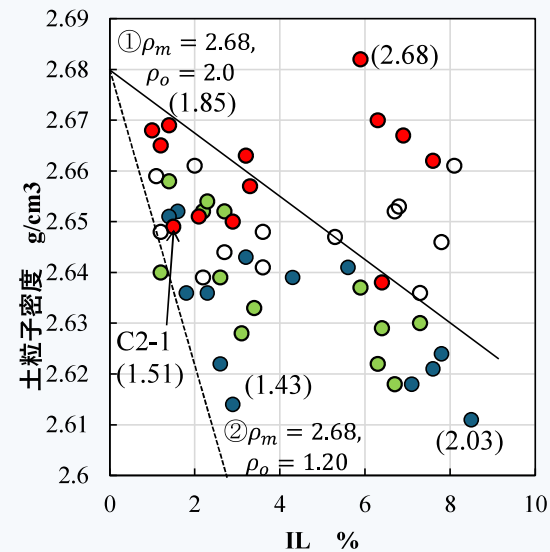
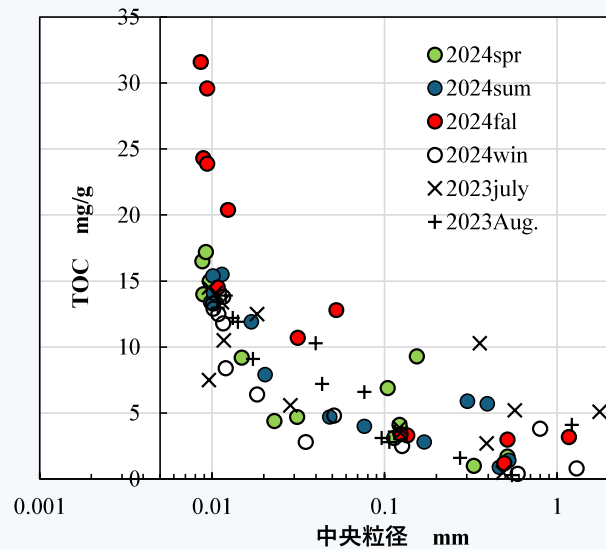
→ 粒径の大きい有機物は生物の生存指標

有機物 (TOC)/細粒分は細粒分に吸着しない有機物の指標となる

→ TOCが少なくてもTOC/細粒分比が砂にTOCが含まれることを示している

軽い有機物と重い粒子の関係をRAG化する (MD5)

死亡直後の底生動物は錯体化（有機物に金属Fe、Sが配位）が進行し重くなる



有機物は分解して細粒化（20μm程度）→ 同粒径の粘土に吸着する
このため土粒子密度を2.68g/cm³と仮定、IL～土粒子密度関係は吸着有機物の分解度の指標とする

AIに使わせるために登録すべき知識カード（MD）

ポイント：熟練者の「見方」をカード（MD）化し、検索される形にしておくと、AIは調査データごとに同じ判断軸を使える。

元素比・BOMカード

Redfield比、C/IL $\approx$ 0.4、新生BOMの目安を整理

強熱減量・燃焼温度カード

ILが代表する成分、燃焼温度による比較条件を明示

C/IL判定カード

0.4付近、0.2付近、0.1以下などの解釈ルール

粒径・細粒分カード

D50・細粒分から砂場/泥場・有機物集積場を判定

土粒子密度カード

有機物混入、無機物結合、難分解化の推定に使う

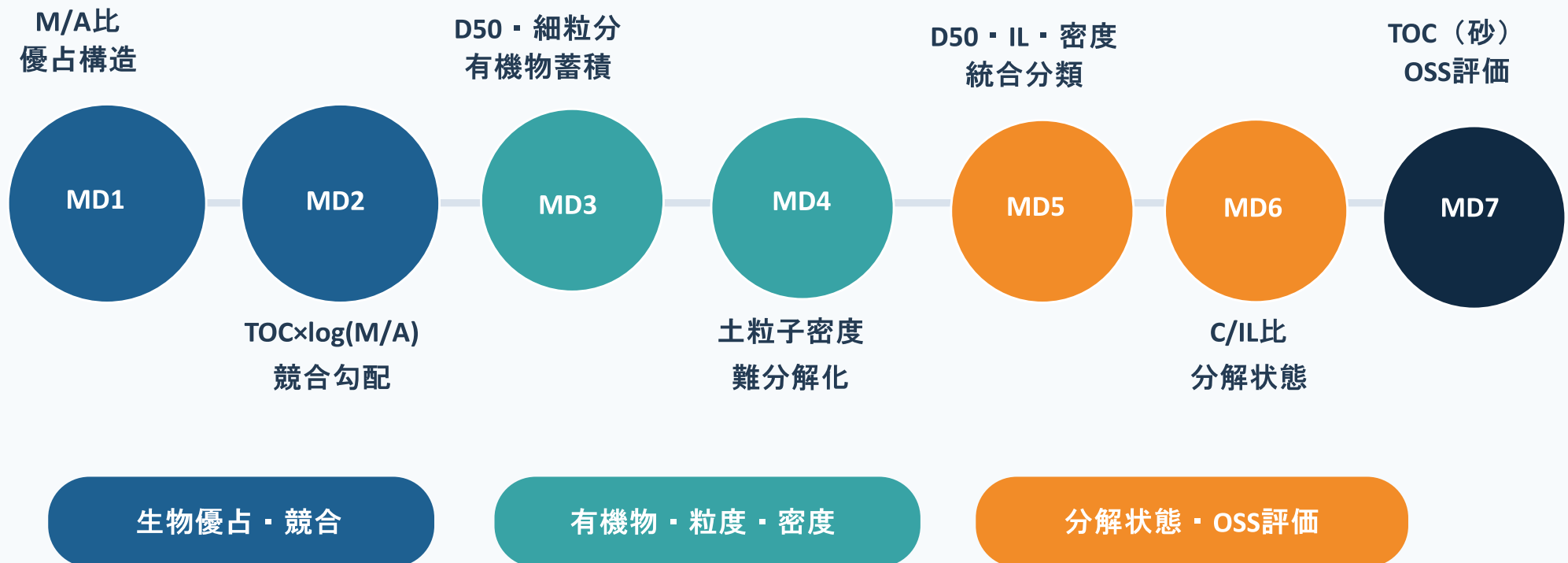
断定禁止・追加分析カード

AIは化学種を断定せず、候補と追加分析を示す

発展形：一般調査員が使いにくい化学的手掛かりを、DB化してAIの標準判断にする

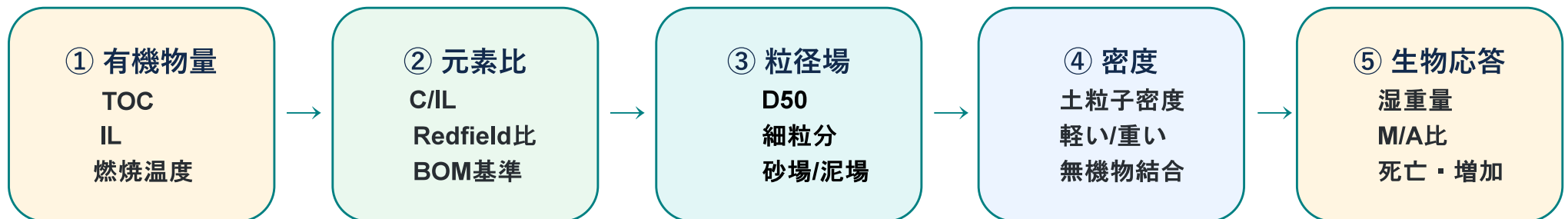
知識DBの中身：MD1～MD7、沿岸環境調査データから技術者が経験に基づき7つのMDを作成

MDは単なる説明文ではなく、「どの分解状態を見て、どのように生物・底質を読むか」の判断基準

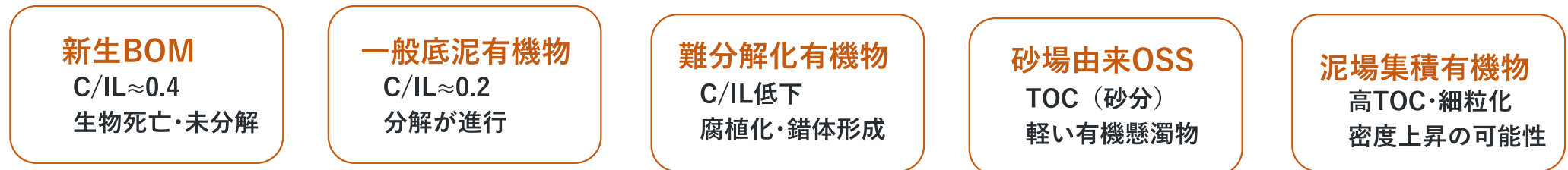


泥内有機物データを読むための判断フロー（ルーティング）がRAG

目的：TOCの大小だけでなく、元素比・燃焼特性・粒径・密度を組み合わせ、有機物の由来と変質段階を推定する。

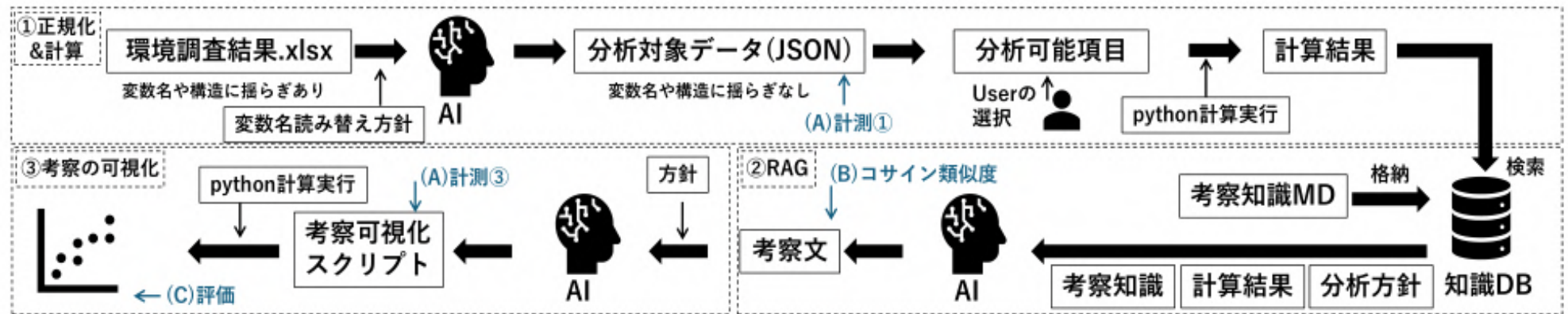


推定される有機物の状態・由来



重要：AIの役割は「化学種の断定」ではなく、複数指標に基づく候補と機構の推定

処理フロー：AI・RAG・Pythonの役割分担



処理フローの概要（青字は性能検証）

AI
表構造の揺らぎ吸収
変数名の読み替え
考察文・スクリプト生成

Python
確定的な計算
JSON生成
安全な実行

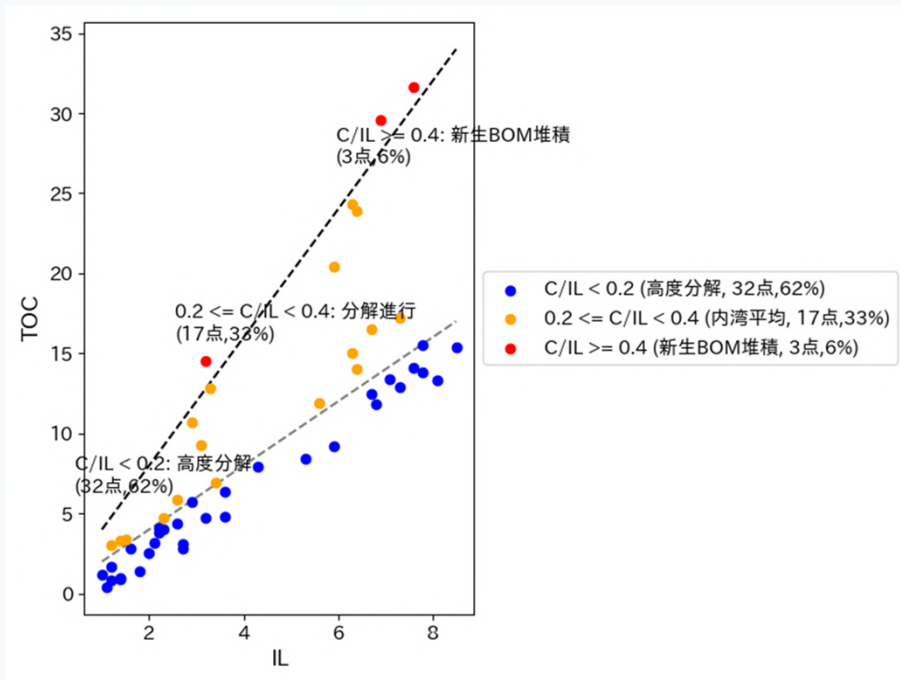
RAG
分析対象に適した
考察知識MDを検索
結果へ適用

評価上の重要点：柔軟性はAI、再現性はPython、判断基準はRAGに分担させている

JSON（JavaScript Object Nation）：エクセルをAIに読みやすくする

AIによる作成グラフの評価①：IL～TOC関係を優先的に表示

分解に伴って元素の存在比が変化



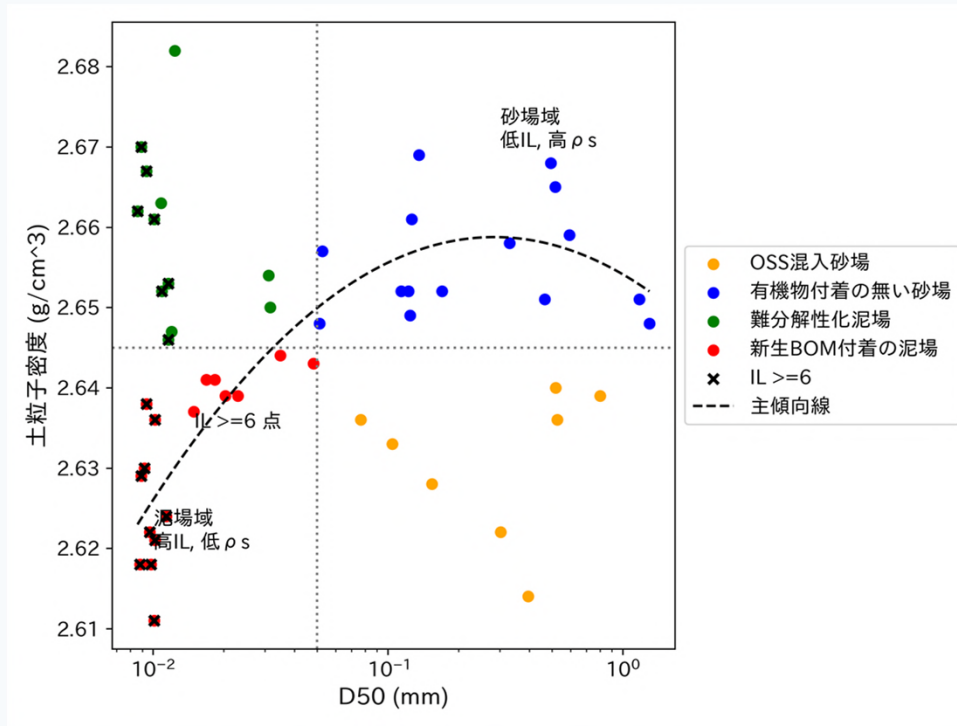
RAGチェック

DBは知識の保管庫、RAGは必要な知識を探してLLMに渡す、LLMはRAGに沿ってグラフを生成できている

C/IL比の意味が補助線として組み込まれたC/ILの意味を区分（考察）して凡例（分布状況）を独自に作成している

評価：データ入力～考察文生成までは実証段階として十分有効

作成グラフ評価②：D₅₀～土粒子密度関係



RAGチェック

与えた知識を補助線として描画できている
グラフ描画ルールをRAGに構造化して保管

分類と注釈文を生成して記述
外れ値を明確に分類

グラフ描画ルールをRAGで考察知識を与えればAIはその知識をグラフに反映できる

最終評価

本研究のRAGは、文書を探して答えるためのRAGではなく、技術者の判断知識をMD・DB化し、観測データとPython計算結果に適用する「考察知識適用型RAG」である。

1.技術者の見方を知識化

M/A比、C/IL比、D50、IL、土粒子密度、TOC(砂分)などを、底生動物・底質環境を読むための判断基準としてMD化した。

2.AIの考察を外部知識に基づかせる

RAGにより、AI単体では見落としやすい元素比、強熱減量、燃焼温度、有機物の分解状態などを毎回参照できるようにした。

3.複雑な考察を図に反映できた

C/IL比による有機物分解状態の区分や、D50～土粒子密度関係に基づく分類など、RAGで与えた知識を補助線・凡例・注釈としてグラフに反映できた。

技術者の経験的な判断知識をDB化し、RAGによりAIへ適用することで、沿岸環境・底質・生物調査データを、再現性のある考察と分類図に変換できる。