

# 高精度大規模**RAG**エンジン ChatBrid の技術的優位性

- LLMは大規模化で精度向上するがRAGは通常精度低下
- 様々な技術的、内容的工夫で遂にナレッジマネジメントの夢が実現！

# 独自のベクトルストレージとTF\*IDFの重み調整

- 独自のベクトルストレージにより、TF\*IDFの重みを自在に調整
- 日本語のSTOP WORDS (不要語: の、こと、もの、する 等の重みをゼロに)を数百語登録
- 単語分割を精緻化。複合語全体と構成語を共にベクトルの要素とし検索精度を向上
  - 沖縄基地問題 vs 沖縄米軍問題 が 67%類似
  - 3.3mg vs 3月3日 が類似しない

(text:型)^3.0 text:溶解 (text:倉庫)^2.0 text:取引  
text:t (text:除く)^2.0 (text:手形)^13.0 (text:日)^15.0  
text:ラード text:鉛 (text:買)^3.0 text:ビートグフ text:  
段ボール (text:名古屋)^2.0 text:雑貨 (text:鋼  
板)^2.0 text:油脂 text:営業 text:テレフタレート (text:  
製品)^2.0 (text:米)^2.0 (text:卸)^3.0 text:叩 text:天  
(text:ポリエチレン)^2.0 (text:纖維)^2.0 ...

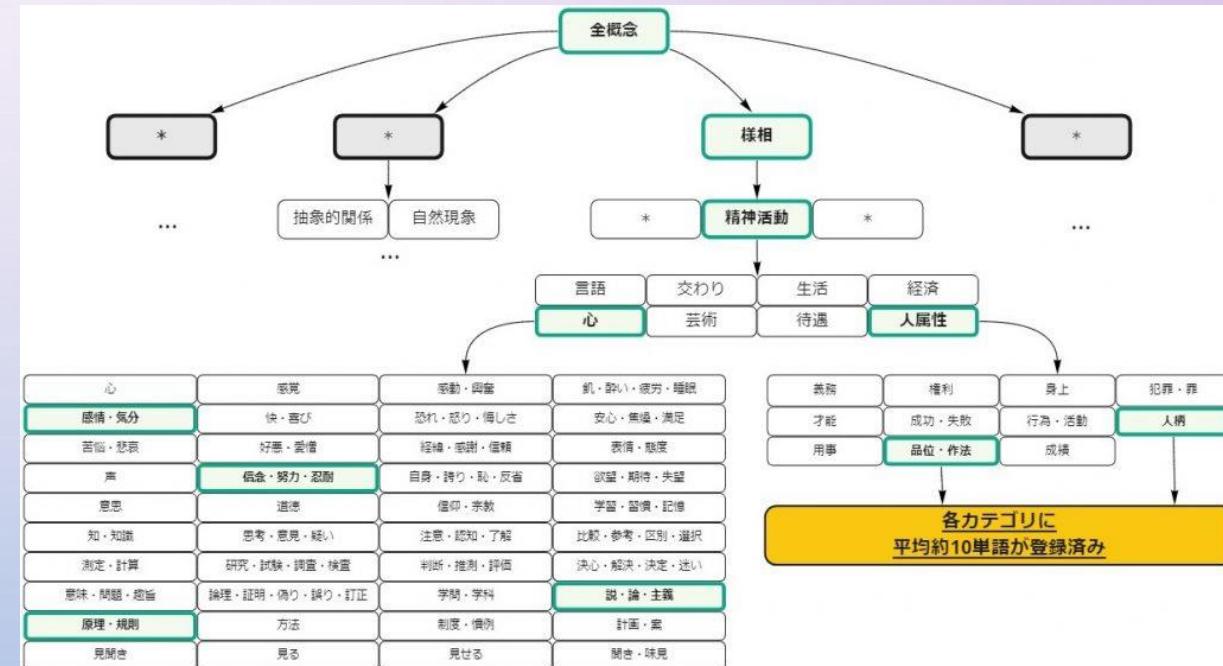


経済特集:日本市場における多様な素材と製品の最新取引動向  
【名古屋発】近年、名古屋を中心とした日本の工業市場では、多岐にわたる素材と製品の取引が活発化しています。特にポリエチレンテレフタレートやポリエチレンなどのプラスチック素材、鉛、油脂といった化学製品の需要が増加しており、これに伴い、営業戦略も急速に変化しています。

最新の市場データによると、名古屋の倉庫ではこれらの素材が大量に取り扱われ、特にポリエチレン製の段ボールやビートグフと呼ばれる特殊な纖維製品の在庫管理が重要視されています。これらの製品は、国内外の雑貨市場での需要が高く、日々の取引量も増加傾向にあります。...

# 5階層シソーラスの活用

- 30数万語の5階層シソーラスを活用し、単語の意味カテゴリをベクトルの要素に追加
- 意味カテゴリをベクトルに追加
- 質問文中に、マニュアルに無い類義語のみが使われている場合でも正しい知識がヒット
  - 一般の質問者は様々な言葉、言い方で表現
  - CHATGPTはそれらの意味を解釈してくれるが、GPT側に送付前の普通のRAGは表現の違いに対応できなかった
  - CHATBRID**なら、程よく類義語に展開可能！
  - 対義語(反対語)も含まれる



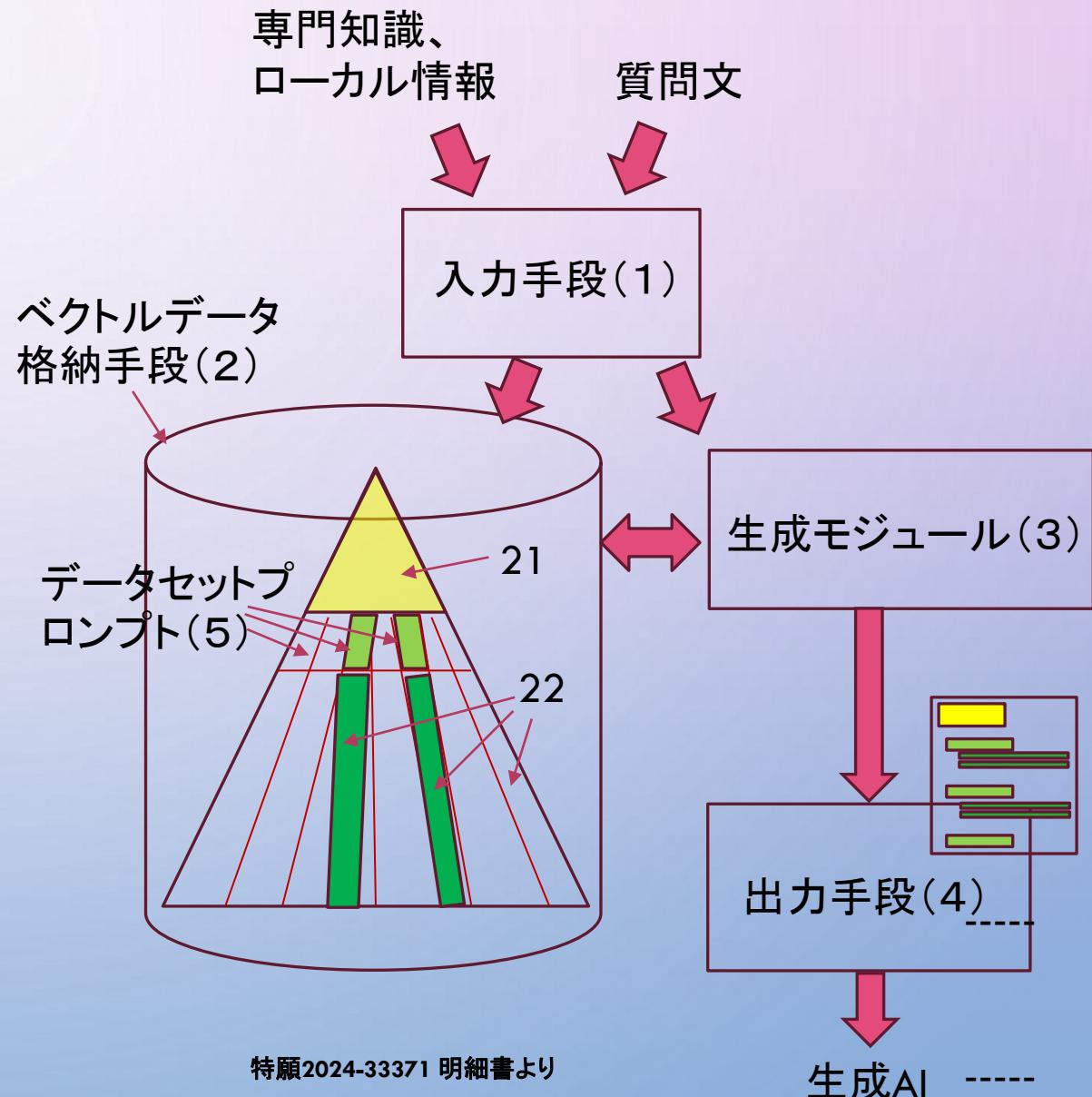
# ベクトル検索結果のビジュアル化 : xAI : explainable AI ～知識デバッグのための視覚フィードバック

- ・自前のベクトルストレージにより、回答不能や不適切な回答のベクトル検索結果を色付け表示
- ・ビジュアル類似検索でAIの回答を秒速で評価  
→ ナレッジ自体の有無、品質を評価、改良
- ・マニュアルの記述が不適切だった場合の即座のビジュアル確認と修正が容易。
- ・マッチした単語と意味コードを強調表示
- ・不適切な回答に対する即時の視覚フィードバック
- ・MARKDOWNをその場で修正 → 即検索に反映

順位	ID	本文	類似度・関連度	グラフ	エージェント名	質問ファイル	行	質問テキスト
1	1103	これまで non-FTE のセキュリティルールは、常駐か非常駐かを基に整理している。コロナ渦で在宅勤務体制が長期化しており、これまでの考え方でセキュリティルールを整備することは不適切。委託業務内容をベースに、non-FTE のセキュリティレベルを定義し直す対象：社員以外全て (amigo、派遣、業務委託、代理店、インター)	248.35896	-	nfte	465		
2	1117	Q&A (追跡追加) これまで non-FTE のセキュリティルールは、常駐か非常駐かを基に整理してた。コロナ渦で在宅勤務体制が長期化しており、これまでの考え方でセキュリティルールを整備することは不適切。委託業務内容をベースに、non-FTE のセキュリティレベルを定義し直す対象：社員以外全て (amigo、派遣、業務委託、代理店、インター)	243.80168	-	nfte	479		
3	609	non-FTE のセキュリティレベル定義は、従来のセキュリティルールと異なる点がいくつかあります。まず、non-FTE のセキュリティレベル定義は、FTE (Full-Time-Equivalent) に比べて、アクセス権限やセキュリティ対策の面で厳しい要件があります。これは、non-FTE が不在の期間が長いため、セキュリティ上の脅威に対してより高いレベルの防護が必要とされるためです。また、non-FTE のセキュリティレベル定義では、データの取り扱いに関するルールが強化されています。具体的には、非公開情報や個人情報などの機密性の高いデータの取り扱いに関する要件が厳しくなります。さらに、non-FTE のセキュリティレベル定義では、社内規定や法令に基づくセキュリティ対策が求められます。これにより、企業や組織が定めるセキュリティポリシーや、国が定める法令に適合することが求めとなります。以上のように、non-FTE のセキュリティレベル定義は、従来のセキュリティルールと比べて、より厳しい要件が求められる点がいくつかあります	148.68065	latest	q2-180	151		
4	610	non-FTE のセキュリティレベルは、主に以下の要素を基準に定義されています。1. アクセス権限: non-FTE がアクセス可能なデータやシステムについて、どのような権限を与えるかが重要な要素の一つです。例えば、非公開情報や個人情報にアクセスする際は、より厳しい権限制御が求められます。2. セキュリティ対策: non-FTE が利用する環境や通信環境、セキュリティ対策などが、セキュリティレベルの決定要素の一つです。例えば、セキュリティソフトウェアの導入や、パスワードの強度によって、セキュリティレベルが異なる場合があります	137.80426	latest	q2-180	152		

# 特許技術による影プロンプト送出量の削減&高精度化

- ナレッジをジャンル分けし各データセットに分割登録することにより影プロンプトの平均量を削減。
  - RAG内の検索でヒットしたデータセットに付随する影プロンプトのみをLLMに送信。
  - これによりLLMのレスポンスが「速い」「安い」「高い(高精度)」
    - 高速な応答と低成本のLLM利用
    - 大規模化しても高精度を維持
    - 通常1専門分野に特化するRAGをマルチ専門家に
      - 質問文に関連の深い上位ナレッジに付随のデータセットプロンプトのみをLLMに送出
      - コンシェルジェ、総務的な機能が可能に
    - 数百本、数千本のマニュアル、数万ページ、数10万件のナレッジを取り入れられる。



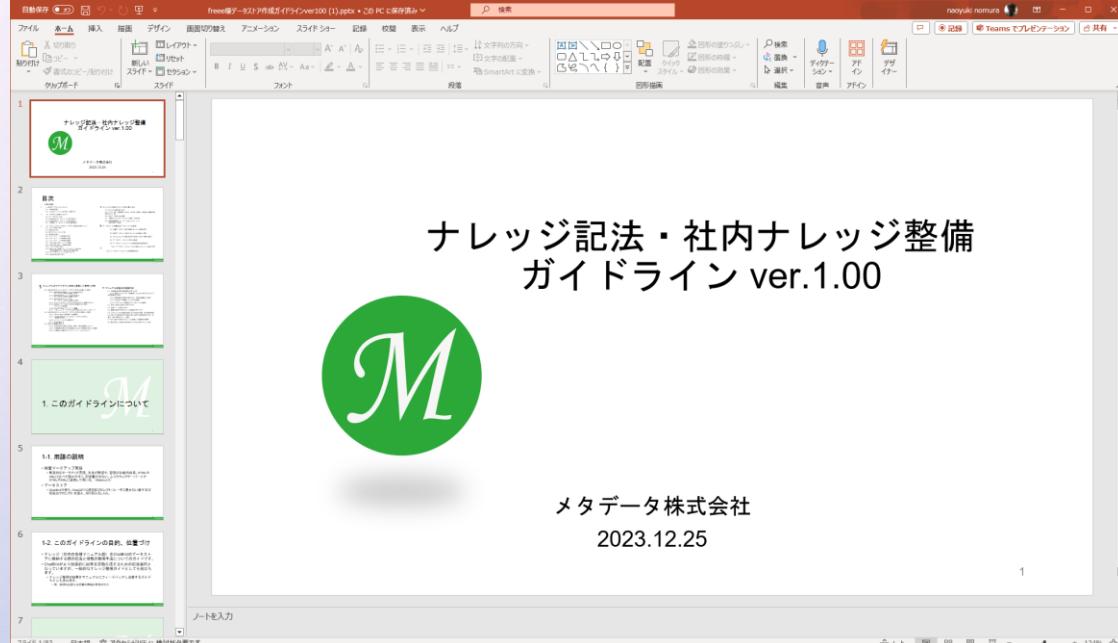
## データセットや各種絞込み条件下の単語分布を目視

- ・高頻度の未知語から、用語説明を付すべき社内専門用語などを素早く発見



# 知識管理のための包括的なマニュアル100数10ページ

- DOCX, XLSX, PPTXをLLMが理解しやすいマークダウンに変換し、知識管理を効率化。
- 100ページ超の効率的なMARKDOWN変換マニュアル
- DOCX、XLSX、PPTXフォーマットに対応



ナレッジ記法・社内ナレッジ整備  
ガイドライン ver.1.00

メタデータ株式会社  
2023.12.25

## 目次

用語の説明	
1. このガイドラインについて	4. マニュアルをMrテキスト分析に取り込み
1-1. 用語の説明	4-1. マニュアル類の取り込み
1-2. このガイドラインの目的、位置づけ	4-2. データ一覧・分類画面、および、ID=2の「本文」に含まれる単語の意味カテゴリ一覧
2. マークダウン形式について	4-3. CSVデータ取り込み画面
2-1. マークダウンとは	4-4. 「単語ランキング」デフォルト画面（上位100）
2-2. CHATBRIDのデータストアで使う記法1	4-5. 表示語数調整スライドバーを右へスライドして上位201件、400件
2-3. CHATBRIDのデータストアで使う記法2	5. データセットの構成法とプロンプトの拡充
2-4. 【参考】データストアでは不要な記法	5-1. 目標データセット数と内部のマニュアル数の目安
3. データストアとしてのマークダウン記法のポイント	5-2. 新規データセット作成とマニュアルの追加、削除
3-1. リストの取り扱い	5-3. マニュアルごとに異なる区切り文字とそのメモ欄への記入
3-2. 表の取り扱い	5-4. データセットプロンプト記入の基本
3-3. 表をテキストにした例	5-5. データセットプロンプトへの用語説明の効率的方法
3-4. 図の取り扱い	5-5-1. データセットプロンプトから全体プロンプトへの格上げ条件
3-5. フローチャートの文章化方法	5-5-2. データセットプロンプトの効果的活用法
3-6. フローチャートの文章化の例	
3-7. フローチャートの文章化結果	
3-8. URLの取り扱い（リンク表記）	
3-9. URLの取り扱い（画像の表示）	
3-10. テストの取り扱い	
3-11. 試行により判明したデータストア内の用語の調整TIPS: 発見方法を追記予定	
3-12. マークダウンファイル名について	
3-13. Q&A形式の取り扱い	
4. マニュアルをMrテキスト分析に取り込み	
4-1. マニュアル類の取り込み	6. マニュアルをマークダウン形式に変換した事例と注意点
4-2. データ一覧・分類画面、および、ID=2の「本文」に含まれる単語の意味カテゴリ一覧	6-1. PPTX形式のマニュアルをマークダウン形式に変換した事例
4-3. CSVデータ取り込み画面	6-1-1. 請求書作成依頼マニュアル（PPTX形式）をマークダウン形式に変換した例
4-4. 「単語ランキング」デフォルト画面（上位100）	6-1-2. 顧客申請マニュアル（PPTX形式）をマークダウン形式に変換した例
4-5. 表示語数調整スライドバーを右へスライドして上位201件、400件	6-1-3. PPTX形式のマニュアルをマークダウン形式に変換する流れ
	6-1-4. マニュアルをマークダウン形式に上手く変換するコツ
	6-1-5. 効率良くマークダウン形式に変換するための、マニュアル改善案
	6-1-6. POWERPOINTのアウトライン編集
	6-1-7. 「今ここ」マーク付きで目次が繰り返し出てくるタイプ
	6-2. DOCX形式のマニュアルをマークダウン形式に変換した事例
	6-2-1. 「人事制度」マニュアルの変換例
	6-2-2. DOCX形式のマニュアルをマークダウン形式に変換する流れ
	6-2-3. ハイパーリンクを変換する
	6-2-4. 表を変換する
	6-3. 見出しの階層を整える
	6-3-1. CHATBRIDが分割する見出し単位、概ね7種類について
	6-3-2. 分割結果が概ね1800字未満になるよう最深の見出しきを調整
	6-3-3. 分割部分を極力コンテキストフリーにすべきこと
6. マニュアルをマークダウン形式に変換した事例と注意点	
7. マニュアル改善点の指摘方法	
7-1. 用語解説対象の候補単語の見つけ方	8. データセット、プロンプト類のバックアップ、移設
7-1-1. 単語ランキングで「定義文」を入れたほうがよさそうな高頻度語を探す	8-1. CSV出力とインポート
7-1-2. 認似検索で単語の使われ方、文脈の難解さを評価	8-1-1. CSV出力対象ファイルについて
7-1-3. ChatGPTが理解しているかを確認	8-1-2. 取り込み時の注意～タイムスタンプ、データ形式等
7-1-4. いがどこにも定義されていないことの確認	8-1-3. CSVをExcel等で編集する際の注意
7-2. 見出し語の不適切の見つけ方	8-1-4. CSVインポート時のエラー、修正方法等
7-3. 欠落ページの見つけ方	8-2. プロンプトの手動バックアップとその対象
7-4. 複数の意味で使われている言葉の見つけ方	8-3. パラメータ設定のバックアップ
7-5. 元マニュアルの体裁を整えるための空白欄、改行類の削除	8-4. ユーザ辞書のバックアップ
7-6. 明らかな漢字脱字や文脈に継ぎ過ぎた省略表現などは、対象を一意に明示すべく、補筆	9. マークダウン化3rd partyツール、Office365からの変換
7-7. GPT内部で英訳されることを考慮した曖昧性の排除	9-1. オンラインのマークダウン編集ツールNotePM
7-8. 最小見出しを概ね1800字以下にする分割テクニック例	9-2. MS Office文書からの変換
	9-3. xdoc2txtの活用
	9-4. ChatBrid上で編集。マスター管理する際の留意点～メモ欄の活用

# 自動レコード分割

## ～マークダウンの見出し階層指定に基づきチャンキング

# 職務発明規定  
《社外秘》メタデータ株式会社

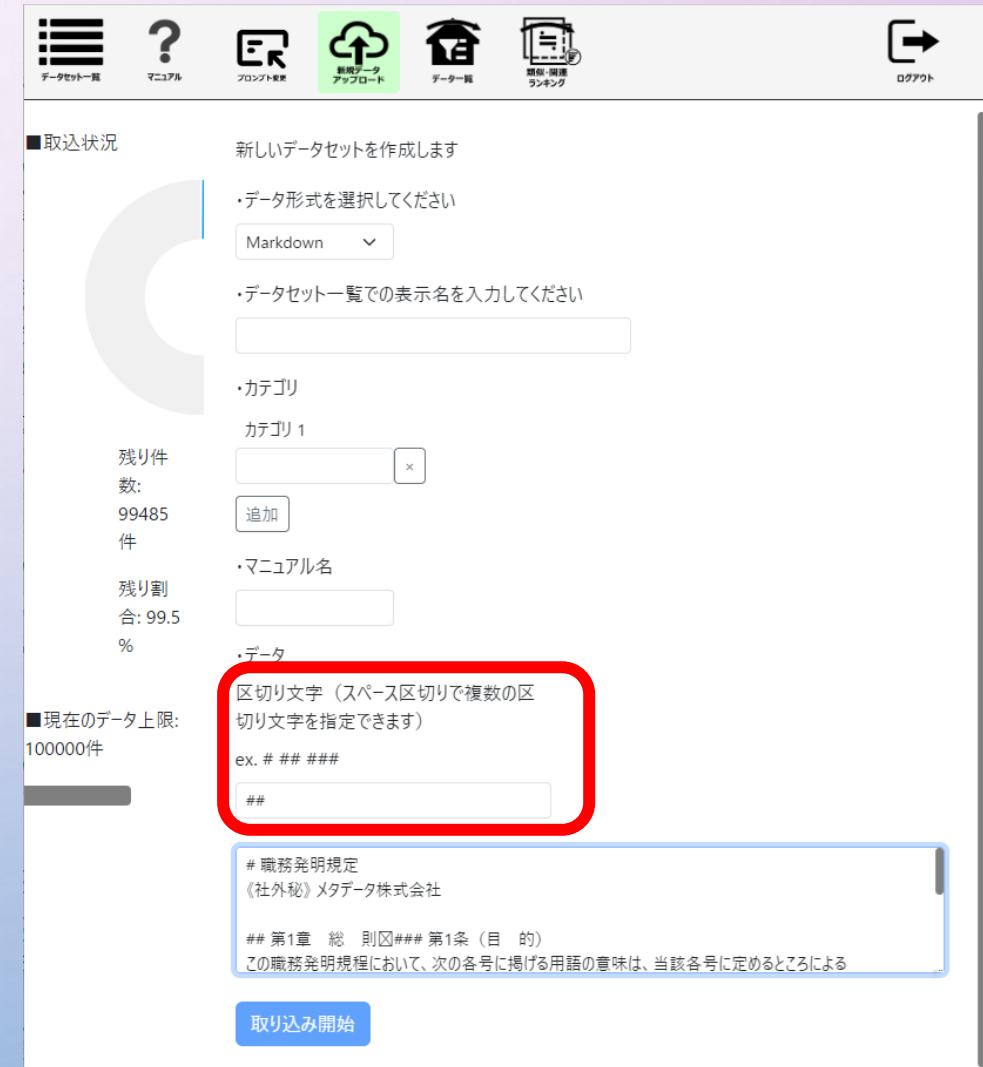
## 第1章 総 則

### 第1条（目的）

- この職務発明規程において、次の各号に掲げる用語の意味は、当該各号に定めるところによる

### 第2条（用語の定義）

- この職務発明規程において、次の各号に掲げる用語の意味は、当該各号に定めるところによる。
  - 職務発明 その性質上会社の業務範囲に属し、かつ、その発明をするに至った行為が会社における従業者等の現在または過去の職務に属する発明として第5条に基づいて会社が認定したもの。
  - 発明者 発明をした従業者等。
  - 従業者等 期間の定めの有無を問わず会社が雇用する者および会社の役員。



■取込状況

新しいデータセットを作成します

・データ形式を選択してください

Markdown

・データセット一覧での表示名を入力してください

・カテゴリ

カテゴリ 1

残り件数: 99485 件

追加

・マニュアル名

残り割合: 99.5 %

■現在のデータ上限: 100000 件

区切り文字（スペース区切りで複数の区切り文字を指定できます）

ex. # ## ###

##

# 職務発明規定  
《社外秘》メタデータ株式会社

## 第1章 総 則

### 第1条（目的）

この職務発明規程において、次の各号に掲げる用語の意味は、当該各号に定めるところによる

取り込み開始

# 知識粒度を小さめにして高精度化しつつコンテキスト情報(上位見出し)を自動付加しLLMの理解を促進

- 知識粒度(レコードサイズ=文字量)がバラつき過ぎると精度低下
- 特に巨大レコードには複数トピックが存在し質問文に無関係なのに関係ありげな記述が混入し精度低下
- 個々のレコードは1パラグラフ程度が好みがコンテキスト不明になってLLMが理解失敗することも



- 小粒度でのコンテキスト維持のために上位見出しを自動付加！

# EXACT MATCH機能と属性による知識絞り込み (近日リリース)

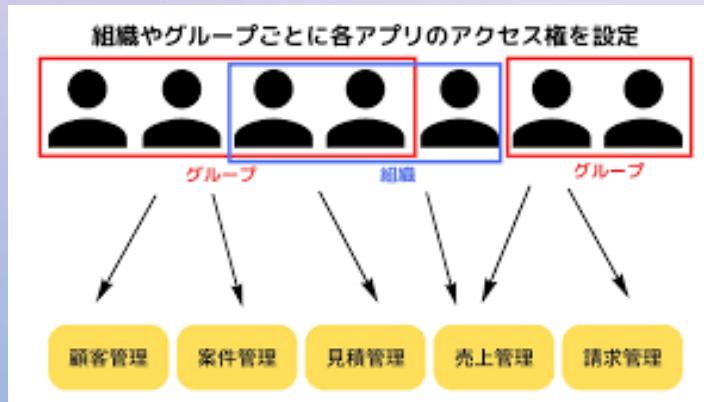
- 指定した文字列とのEXACT MATCHを条件にすることで材料名の取り違えなどの致命的な誤りを防止。
- 属性、法律・条令の有効期間、適用地域などによる知識の厳密な絞り込みへ。

材料ID	材料名	純度	供給元	在庫量 (kg)	備考
Cu-101	電気銅 (C1100)	99.99%	A社	1000	高純度銅
Cu-102	りん脱酸銅 (C1220)	99.90%	B社	500	少量のりんを含む
Cu-103	無酸素銅 (C1020)	99.96%	C社	750	酸素含有量が極めて低い
Cu-104	タフピッチ銅 (C1100)	99.90%	D社	1200	一般的な純銅
Zn-201	高純度亜鉛 (Z1100)	99.99%	E社	600	高純度亜鉛
Sn-301	高純度スズ (Sn99.9)	99.90%	F社	400	高純度スズ
Ni-401	電解ニッケル (N02200)	99.90%	G社	300	高純度ニッケル
Al-501	工業用純アルミ (A1050)	99.50%	H社	800	一般的な工業用純アルミ

Cu-101とCu-102やZn-101が類似検索されてはならない。  
Exact matchで絞り込み、該当レコードのデータのみを回答に反映させるべし。

# 知識アクセス権制御 (開発中)

- 質問者の所属や資格を確認し、アクセス権限内のナレッジ参照を実現。
  - セキュアで関連性の高い情報の取得
  - 企業のセキュリティとコンプライアンスに重要



- ・ジャンル別マニュアル群
- ・データセット単位でも、その中のマニュアル単位でも、知識レコード単位でもアクセス権設定可能

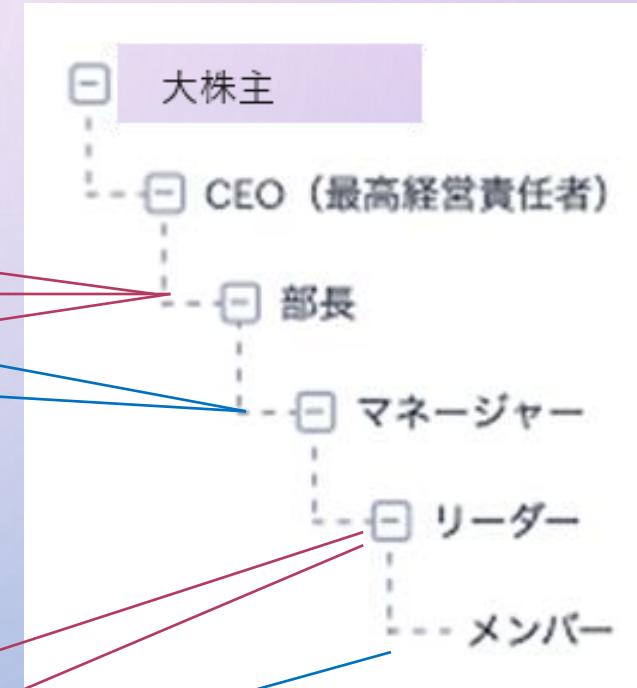
上部画面: データセットのリスト

データセット名	取扱情報	作成日時	最終アクセス日時
契約・顧客向け規約等	開く	2024/02/09 19:30:35	2024/02/19 20:02:21
test	開く	2024/02/05 17:09:44	削除
社内規約類	開く	2024/01/08 15:44:20	削除
Chatbridマニュアル	開く	2023/12/13 08:33:01	削除

下部画面: クリックで表示する条件

条件:

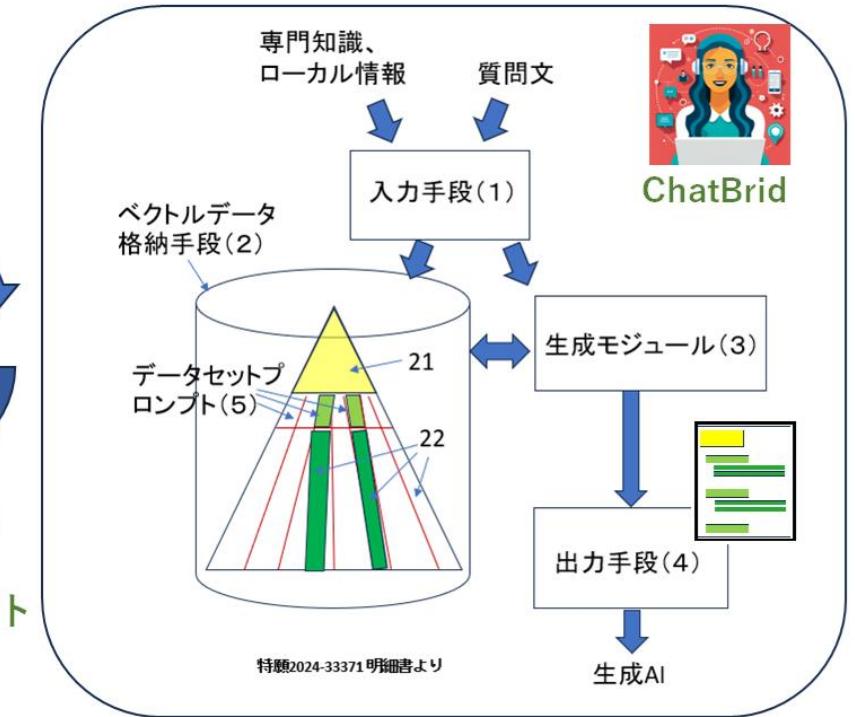
- 契約・顧客向け規約等
- 社内規約類
- Chatbridマニュアル



# APIモードでの統合運用 ～企業システムとの相互通接続性

- CHATBRID全体がAPIモードで動作
- 各種企業情報システムと一体化して運用可能。
- アプリ構築環境MIIBOやDIFYと統合済。
  - シナリオベース、ワークフローベース

## webhookによる連携 (ChatBridのAPI)



# Miibo '内蔵' のChatBridが回答を送出

webhook試作用  
 webhook試作用です。  
 提供元: metadata inc.

Powered by miibo

【プレビュー】  
 御用ですか？

メタデータ株式会社の社員が業務遂行において発明をした場合、どうすればいいですか？

【プレビュー】  
 以下の手順で対応してください。  
 1. 社員は、発明の内容を遅滞なく会社に届け出る必要があります。  
 (職務発明規定第4条)  
 2. 会社は、届出のあった発明が職務発明に該当するかを認定します。  
 職務発明と認定した場合、会社が特許を受ける権利を承継するかどうかを決定します。  
 (職務発明規定第5条第1項)  
 3. 会社が特許を受ける権利の承継を決定した場合、正当な理由がない限り、直ちに特許出願を行います。  
 (職務発明規定第5条第2項)  
 4. 会社が特許出願すると判断した場合、社員は発明の権利の譲渡書に押印または署名をする必要があります。  
 (業務委託契約書第11条第2項の例)  
 5. 特許出願や権利の維持に関する費用は会社が負担します。  
 6. 発明の所有比率は、発明者の数や貢献度を考慮して、会社と社員の協議により決定します。  
 (業務委託契約書第11条第3項、第4項の例)  
 以上が、社員の職務発明に関する一連の流れです。発明の届出を速やかに行い、会社の指示に従って手続きを進めてください。

△AIによる自動生成

自由に入力してください。

▲

# Dify '内蔵'のChatBridが回答を送出