

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

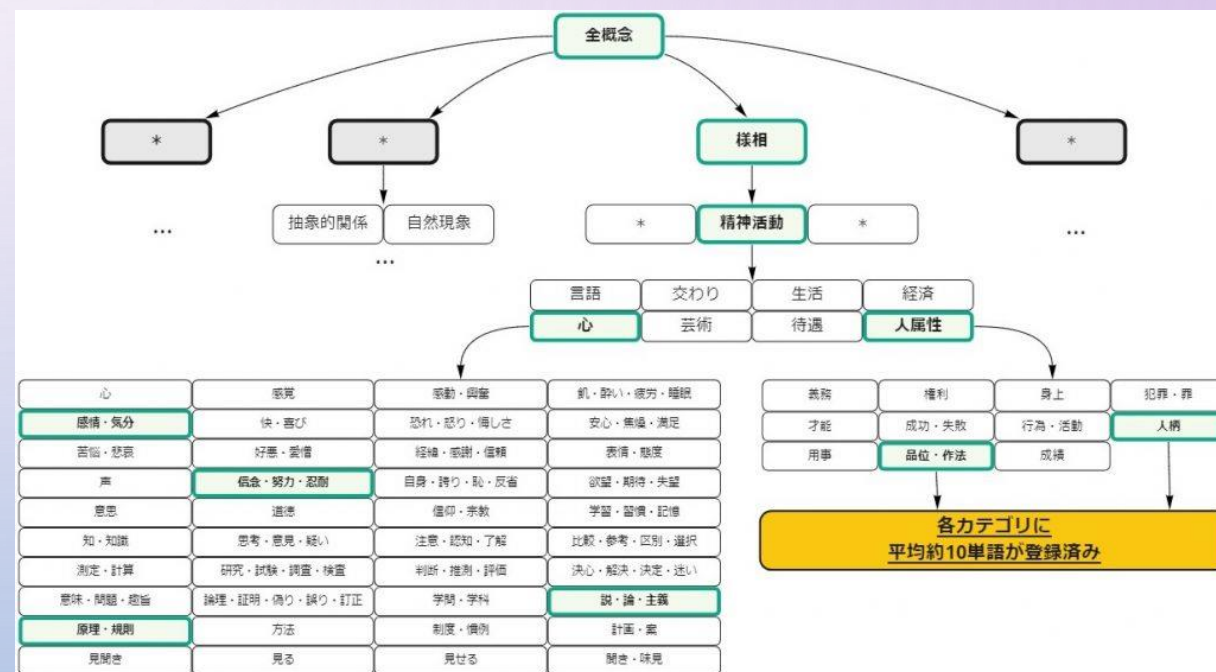
(text:型)^3.0 text:溶解 (text:倉庫)^2.0 text:取引
text:t (text:除く)^2.0 (text:手形)^13.0 (text:日)^15.0
text:ラード text:鉛 (text:買)^3.0 text:ビートグフ text:
段ボール (text:名古屋)^2.0 text:雑貨 (text:鋼
板)^2.0 text:油脂 text:営業 text:テレフタレート (text:
製品)^2.0 (text:米)^2.0 (text:卸)^3.0 text:叩 text:天
(text:ポリエチレン)^2.0 (text:繊維)^2.0 ...



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繊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

The screenshot displays the xAI interface, which is used for visualizing vector search results. The interface is divided into a sidebar on the left and a main content area on the right.

Sidebar (Left):

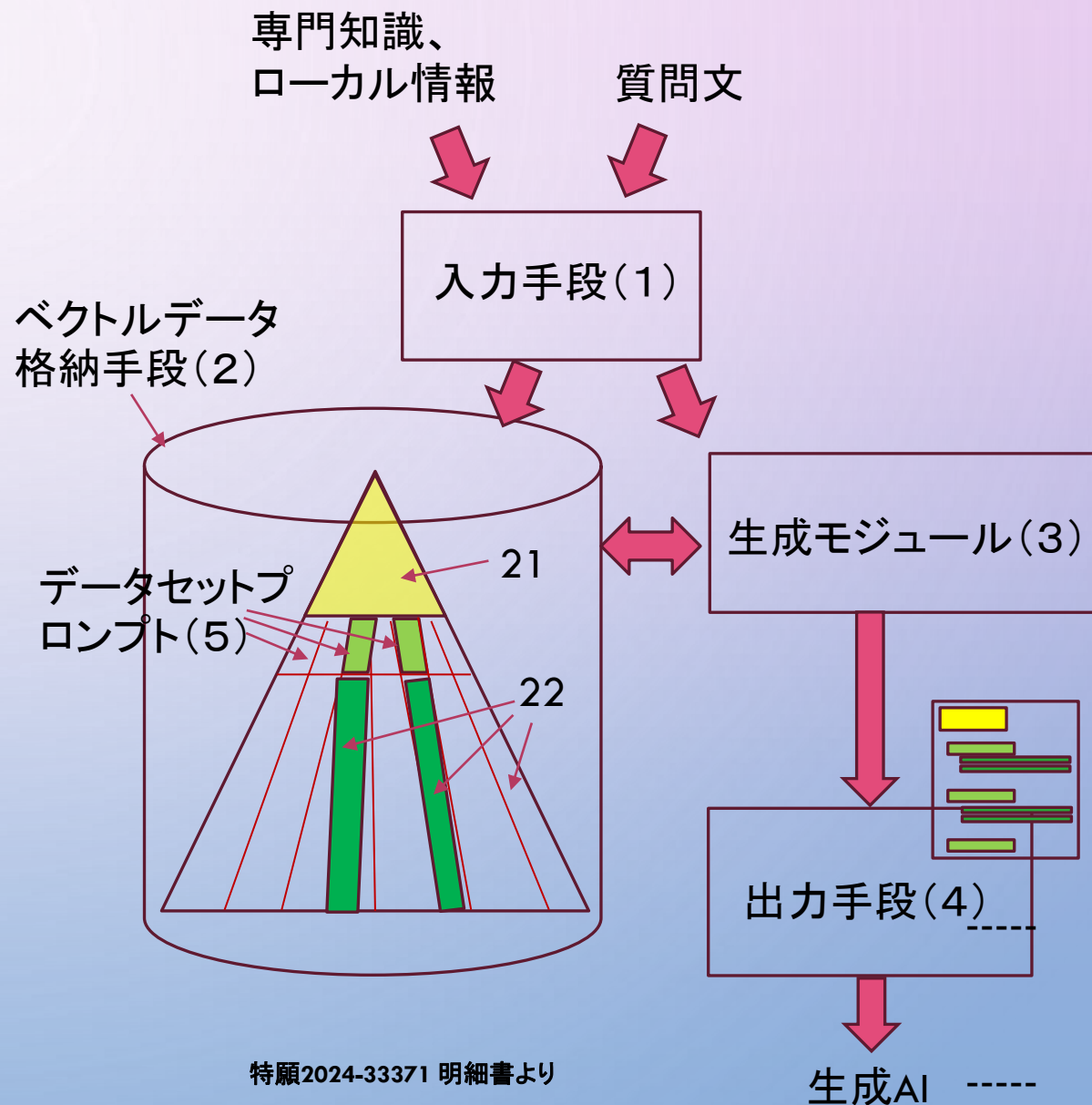
- 検索条件のクリア** (Clear search conditions)
- 検索・絞り込み** (Search & Filter)
- 類似度を文章から検索** (Search similarity from text): A text box containing a paragraph about security rules and FTE (Full-Time Equivalent) levels.
- 検索対象** (Search target): Checkboxes for ☒ 本文 (Main text), ☐ 重要語 (Keywords), ☐ Xモ (X-mo), and ☒ 質問テキスト (Question text).
- 類似度・関連度 正規化** (Similarity/Relevance normalization): A slider set to 0.0 ~ 1.0.
- 最大ヒット件数 (数値)** (Maximum number of hits (numeric)): A dropdown menu set to 100.
- 意味コード検索の重み** (Meaning code search weight): A dropdown menu set to 0%.
- 複集合検索** (Complex search): A toggle switch.
- 文字列** (Text string): A dropdown menu.
- ID (行番号)** (ID (line number)): A dropdown menu.
- ネガボジ** (Negaboji): A dropdown menu.

Main Content Area (Right):

- ファイル名 = 458+707=[q2...]_202306190723**
- 検索文字列 =** `これ | non | fte | セキュリティ | ルール | 常駐 | 非常 | 駐 | 整理 | する | いる | コロナ | 渦 | 在宅勤務 | 体制 | 長期 | 化 | する | こと | 不適切 | 委託 | 業務 | 内容 | ベース | non | fte | セキュリティ | レベル | 定義 | する | 直す | 対象 | 社員 | 以外 | 全て | amigo | 派遣 | 業務委託 | 代理店 | インターン`
- 意味コード検索の重み = 0%**
- 100 件ヒット (全1165件中) *最大ヒット件数が100件に絞られています。**
- 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**
- 50** (dropdown)
- 1 2 >** (pagination)
- | 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 |
|----|------|--|-----------|-----|---------|--------|-----|
| 1 | 1103 | これ non fte のセキュリティルールは、常駐が非常駐かを駐 整理 する いる コロナ 渦 在宅勤務 体制 長期 化 する こと 不適切 委託 業務 内容 ベース non fte セキュリティ レベル 定義 する 直す 対象 社員 以外 全て amigo 派遣 業務委託 代理店 インターン | 248.35896 | | - | nfte | 465 |
| 2 | 1117 | Q&A (随時追加) これ non fte のセキュリティルールは、常駐が非常駐かを駐 整理 する いる コロナ 渦 在宅勤務 体制 長期 化 する こと 不適切 委託 業務 内容 ベース non fte セキュリティ レベル 定義 する 直す 対象 社員 以外 全て amigo 派遣 業務委託 代理店 インターン | 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 いくつがあります。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ので、セキュリティ上の脅威に対してより高いレベルの防御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。 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 することが必要となります。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 いくつがあります | 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 端末や通信環境、セキュリティ対策などが、セキュリティレベルの法定要素の一つです。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」「旨い(高精度)」
 - 高速な応答と低コストの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
 - コンシエルジェ、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

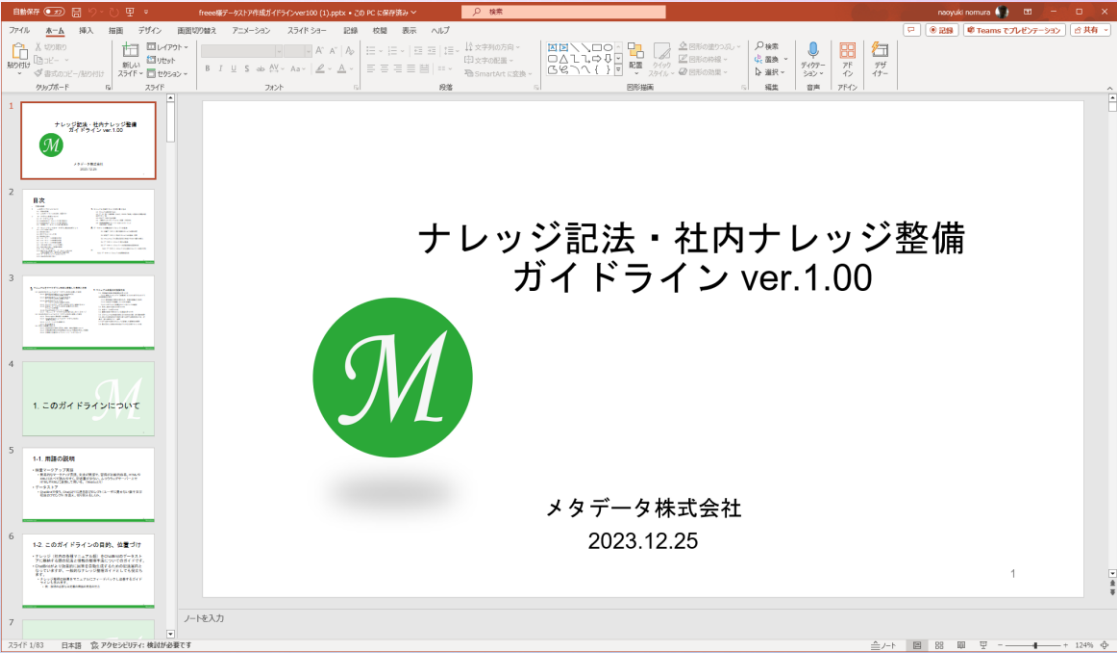


知識管理のための包括的なマニュアル100数10ページ

- 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

目次

1.	用語の説明	4. マニュアルをMrテキスト分析に取り込み	
1.1.	このガイドラインについて	4-1. マニュアル類の取り込み	
1-1.1.	用語の説明	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	
1-2.	このガイドラインの目的、位置づけ	4-3. CSVデータ取り込み画面	
2.	マークダウン形式について	4-4. 「単語ランキング」デフォルト画面' (上位100)	
2-1.	マークダウンとは	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	
2-2.	CHATBRIDのデータストアで使う記法1		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
2-3.	CHATBRIDのデータストアで使う記法2		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
2-4.	【参考】データストアでは不要な記法		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
3.	データストアとしてのマークダウン記法のポイント		6-1-2. 請求書申請マニュアル(PPTX形式)をマークダウン形式に変換した例
3-1.	リストの取り扱い	5-1. 目標データセット数と内部のマニュアル数の目安	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
3-2.	表の取り扱い		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
3-3.	表をテキストにした例	5-2. 新規データセット作成とマニュアルの追加、削除	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
3-4.	図の取り扱い	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	6-1-6. POWERPOINTのアウトライン編集
3-5.	フローチャートの文章化方法	5-4. データセットプロンプト記入の基本	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
3-6.	フローチャートの文章化の例	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	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
3-7.	フローチャートの文章化結果	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	6-2-1. 『人事制度』マニュアルの変換例
3-8.	URLの取り扱い(リンク表記)	5-5-2. データセットプロンプトの効果的活用法	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
3-9.	URLの取り扱い(画像の表示)		6-2-3. ハイパーリンクを変換する
3-10.	テストの取り扱い		6-2-4. 表を変換する
3-11.	試行により判明した、データストア内での用語の調整TIPS:発見方法を追記予定		6-3. 見出しの階層を整える
3-12.	マークダウンファイル名について		6-3-1. CHATBRIDが分割する見出し単位、概ね7種類について
3-13.	Q&A形式の取り扱い		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
			8. データセット、プロンプト類のバックアップ、移設
			8-1. CSV出力とインポート
			8-1-1. CSV出力対象フィールドについて
			8-1-2. 取り込み時の注意 ~タイムスタンプ、データ形式等
			8-1-3. CSVをExcel等で編集する際の注意
			8-1-4. CSVインポート時のエラー、修正方法等
			8-2. プロンプトの手動バックアップとその対象
			8-3. パラメータ設定のバックアップ
			8-4. ユーザ辞書のバックアップ
			9. マークダウン化3rd partyツール、Office365からの変換
			9-1. オンラインのマークダウン編集ツールNotePM
			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。

データセット一覧

マニュアル

フロント編集

新規データアップロード

データ一覧

検索・編集
ランキング

ログアウト

■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

×

追加

マニュアル名

データ

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. # ## ###

##

■現在のデータ上限:

100000件

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ましいがコンテキスト不明になってLLMが理解失敗することもある



- 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！

データセット一覧

検索条件のクリア

検索・絞込

回答

見出し

データ形式

マニュアル名

責任部門

カテゴリ 2

カテゴリ 3

?

マニュアル

データ一覧

データセット
プロンプト変換

条件:
ファイル名 = 社内規約類抄コンテキストフリー化

☐ 折りたたみ

3件 / 全3件中

クリックで表示/非表示を切替

ID 回答 見出し 責任部門 カテゴリ 2 カテゴリ 3 マニュアル名 データ形式 日付時刻 メモ 削除

ID	回答	見出し	責任部門
1		第1章 総則	法務部
2	- この就業規則（以下「規則」という。）は、メタデータ株式会社（以下「会社」という）の秩序を維持し、業務の円滑な運営を期するため、従業員の労働条件、服務規律その他の就業に関する事項を定めるものである。 - この規則に定めのない事項については、労働基準法その他の法令の定めるところによる。 - この就業規則（以下「規則」という。）は、メタデータ株式会社（以下「会社」という）の秩序を維持し、業務の円滑な運営を期するため、従業員の労働条件、服務規律その他の就業に関する事項を定めるものである。 - この規則に定めのない事項については、労働基準法その他の法令の定めるところによる。	第1条 (目的) < 第1章 総則	法務部
3		第2条 (***) < 第1章 総則	法務部

EXACT MATCH機能と属性による知識絞り込み (近日リリース)

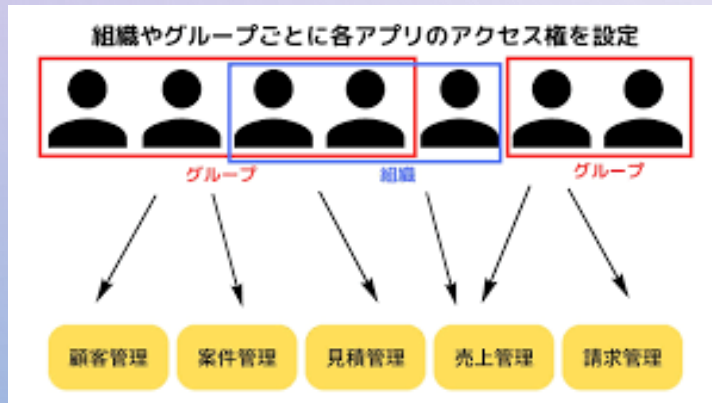
- 指定した文字列とのEXACT MATCHを条件にすることで材料名の取り違えなどの致命的な誤りを防止。
- 属性、法律・条令の有効期間、適用地域などによる知識の厳密な絞り込みへ。

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)

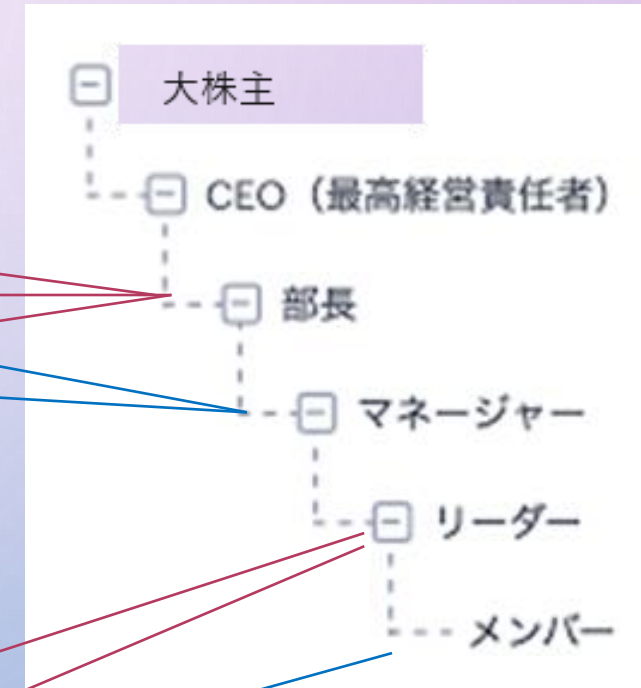
- 質問者の所属や資格を確認し、アクセス権限内でのナレッジ参照を実現。
 - セキュアで関連性の高い情報の取得
 - 企業のセキュリティとコンプライアンスに重要



- 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

The screenshot shows the Chatbird management interface. On the left, there is a sidebar with navigation icons and a progress indicator for 'データセット' (Data Set) showing 99.56% completion. The main area displays a table of data sets with columns for 'データセット名' (Data Set Name), '取込情報' (Import Information), '作成日時' (Creation Date/Time), and '最終アクセス日時' (Last Access Date/Time). The table lists four data sets: '契約・顧客向け規約等', 'test', '社内規約類', and 'Chatbirdマニュアル'. Each row has a yellow '取込' (Import) button and a red '削除' (Delete) button. Arrows indicate the flow of data from the '取込' buttons to the '削除' buttons.

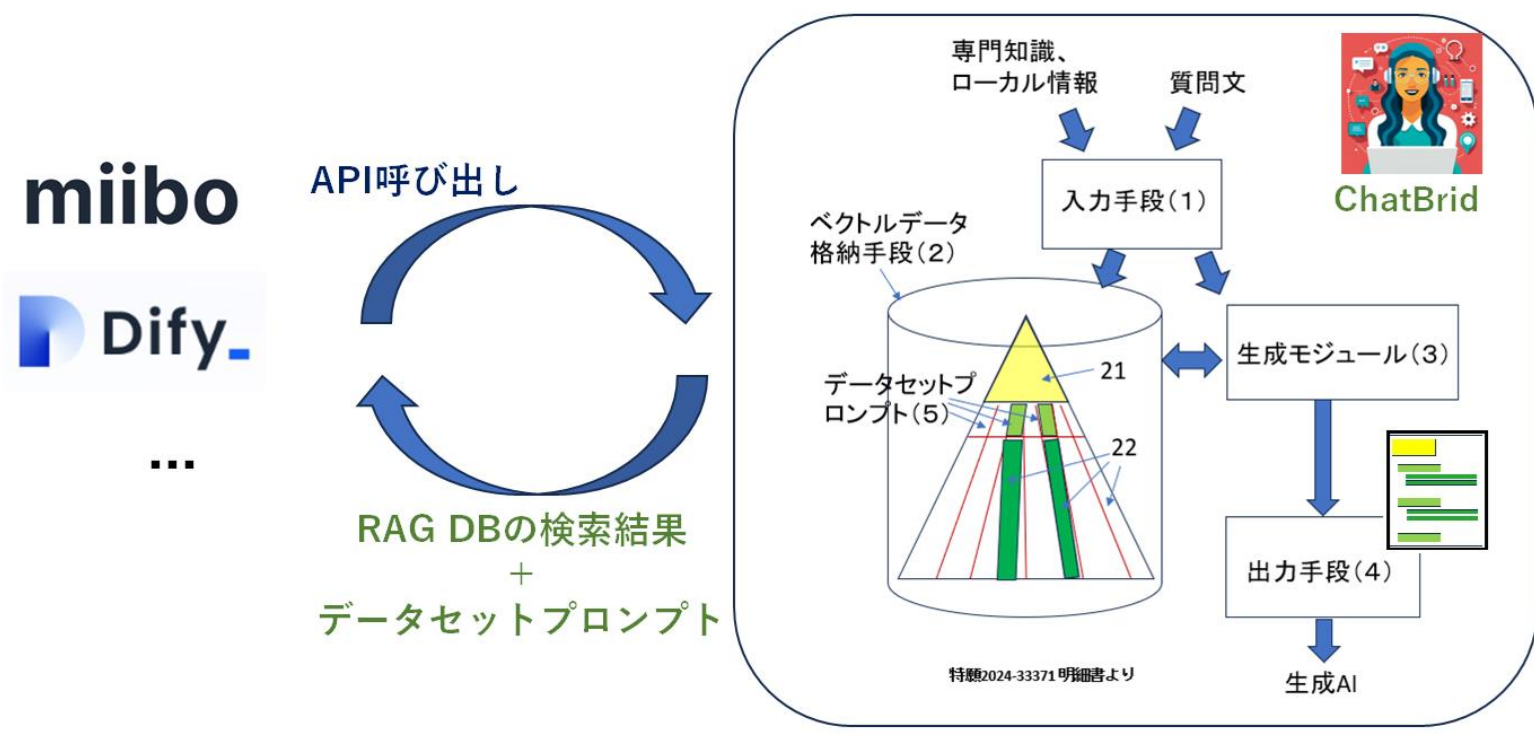
データセット名	取込情報 件数	作成日時 最終アクセス日時	操作
契約・顧客向け規約等	取込済み	2024/02/09 19:30:35	削除
test	取込済み	2024/02/19 20:02:21	削除
社内規約類	取込済み	2024/02/05 17:09:44	削除
社内規約類	取込済み	2024/02/09 20:04:32	削除
社内規約類	取込済み	2024/01/08 15:44:26	削除
社内規約類	取込済み	2024/02/21 12:18:47	削除
Chatbirdマニュアル	取込済み	2023/12/13 08:33:01	削除



APIモードでの統合運用 ～企業システムとの相互接続性

- 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わ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。



Diffy '内蔵'のChatBridが回答を送出

Dify

Star 32,442

探索スタジオ / chatbridapi_chatbotナレッジツールアップグレードD Dify

ログ注釈

ログは、アプリケーションの実行状況を表示します。

過去7日間すべて

時間エンタープライズ

06/04/2024 06:28 AM031ba

会話ログ

詳細トレース

LLM538.880 ms · 1.679K tokens

入力
この文書は、平成18年4月1日から施行する、\n\n容、\n\n就業規則変更履歴、\n\n平成19年5月1日、第15条、第16条、第17条、第28条、第30条に記載。\\n・平成19年7月13日、第29条、2項~4項を19年5月1日に遡って変更。\\n・平成24年1月31日、第14条、第24条、第25条、第26条、第27条追記。\\n28\\n\\n関連度：7.9381765\\n}}"}

3 }

データ処理
1 {
2 "model_mode": "chat",
3 "prompts": [
4 {
5 "role": "system",
6 "text": "Use the following context as your learned knowledge, inside <context></context> XML tags.
\n<n>state": "
}\\transform-as-response\\">"\\n\\noss 名称
出力
1 {
2 "text": "
3 "usage": {
4 "prompt_tokens": 1667,