

Express Finder / シソーラス辞書のご紹介

NTTアドバンステクノロジー株式会社

こんなお困りごとはありませんか？

- 検索システムやFAQサイトで、欲しい情報にたどり着けない…
- データを集計・分析したいが、表現のばらつきを分類するのは大変…



- **日本最大級の同義・類義語電子辞書**で基本語から専門語まで幅広く収録
- **企業名や言葉の表記ゆれ**を解消し、検索システムの**検索ヒット率向上**を実現

point1

検索システムの
検索ヒット率向上

キーワードに対して同義語を含めた検索が可能

⇒検索漏れを防ぎます。

point2

日本最大級の
同義語や類義語が
約**320万語**収録

採用出典を厳選した専門性の高い用語を収集

⇒様々な業界の多岐に渡る検索や問合せに対応できます。

point3

言語の専門家が
定期的に辞書を
最新データに更新

辞書のメンテナンス作業にかかる稼働を削減

⇒余った稼働で他の作業へ注力できます。

Express Finder / シソーラス辞書で解決します

■ 検索システムの検索ヒット率向上や、表記ゆれを解消して集計・分析を効率化



検索システムで欲しい情報にたどり着けない

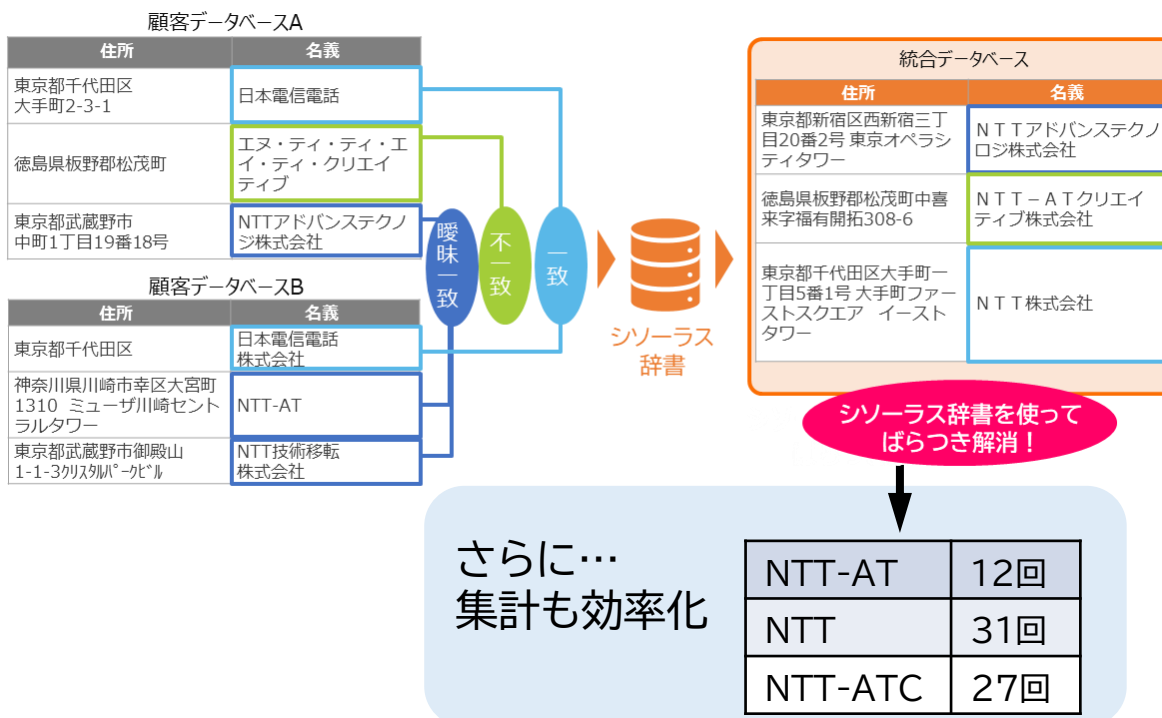
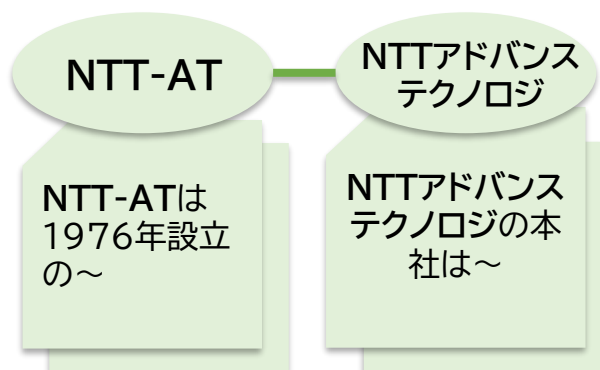


表現のばらつきがあり集計・分析の作業が大変

“NTT-AT”
を検索



↓ 同義語の検索結果

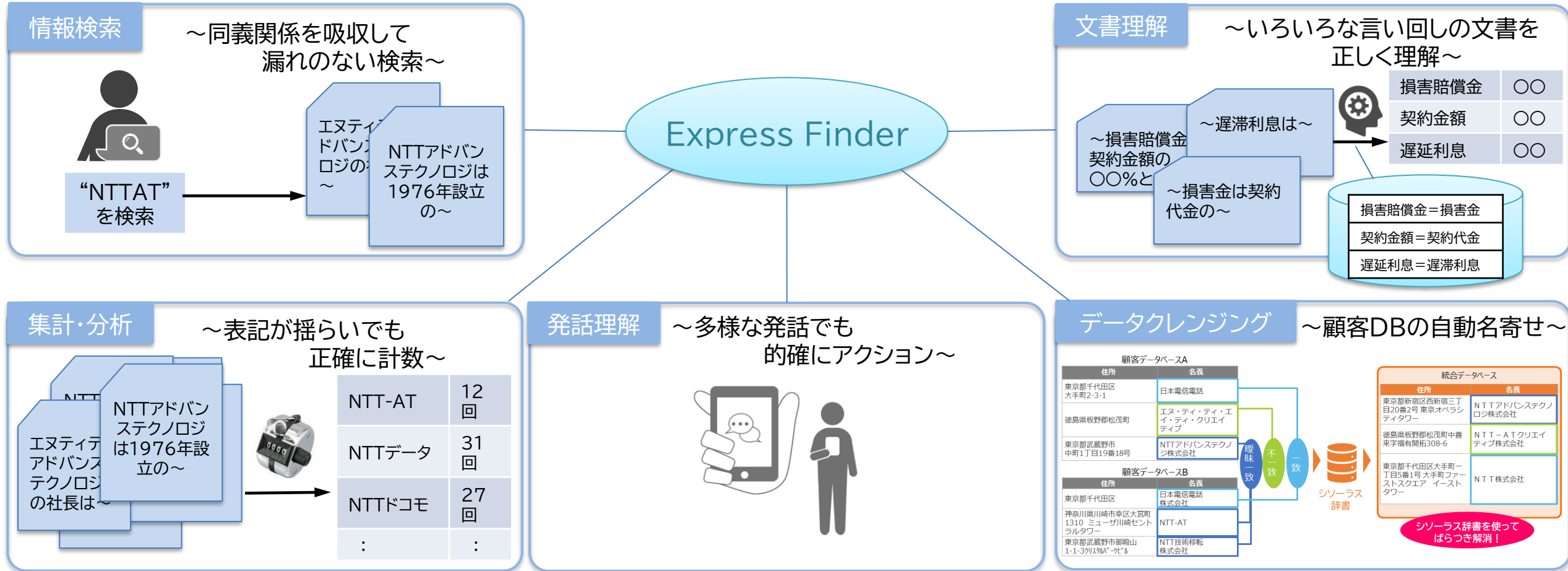


検索ヒット率の向上で検索漏れを防止



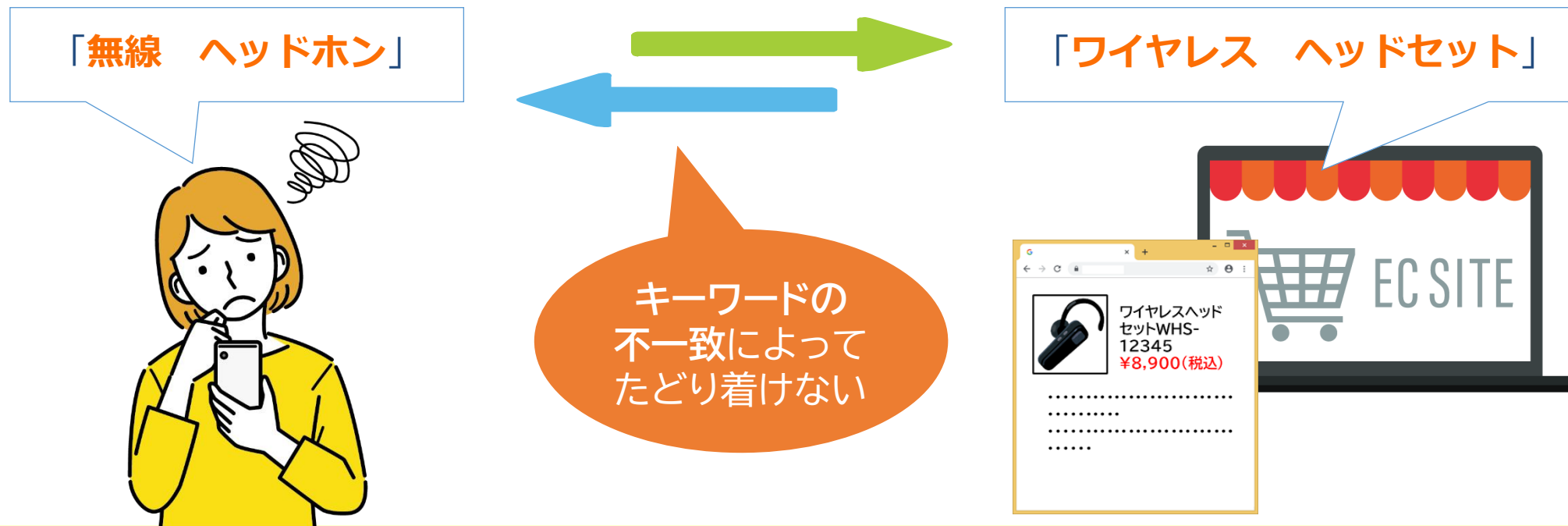
表記ゆれを解消してデータベースを名寄せ
さらに集計・分析も効率化

■ 情報検索、テキスト集計や分析、文書の理解、発話の理解、データクレンジング等のシーンで活用



利用シーン1 | 例えばこんなことはありませんか？

とある通販サイトでの検索例



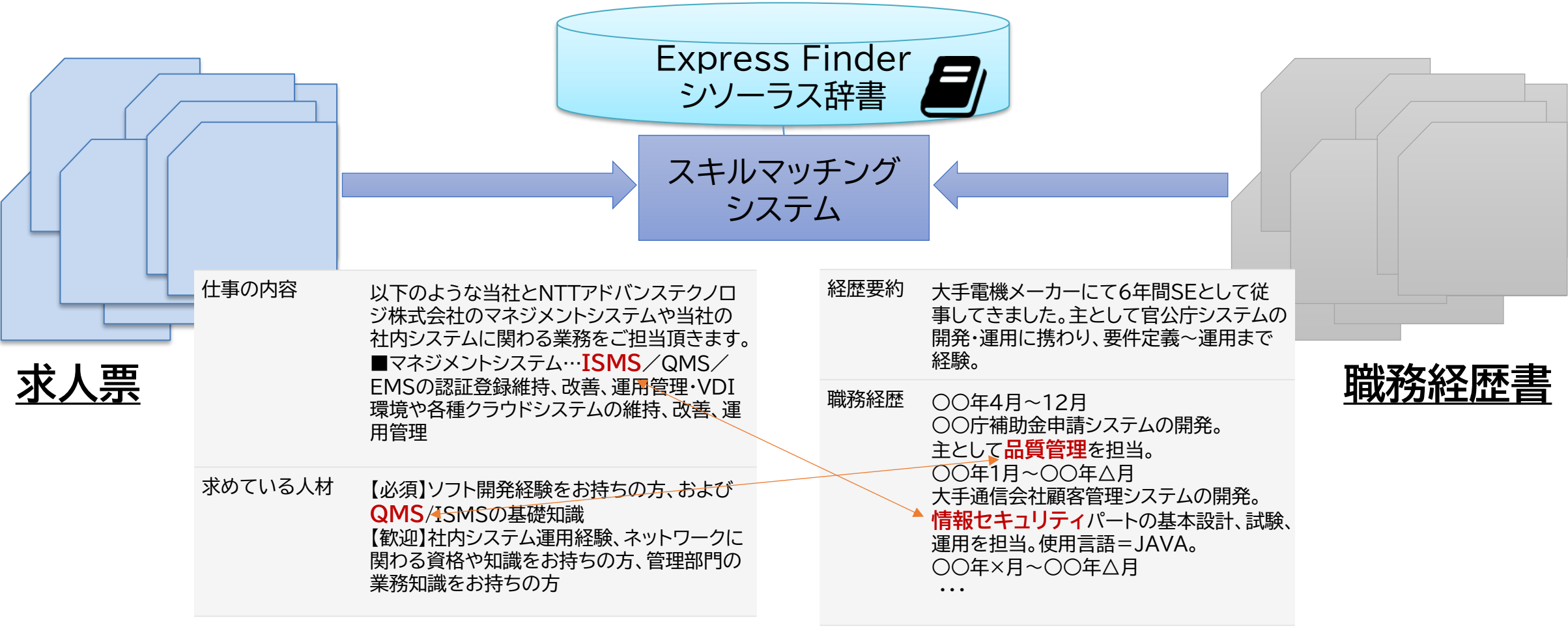
- せっかくの販売サイトなのに、ちょっとしたキーワードの違いで販売機会を逃してしまうかもしれません
- 検索結果から少しずつサイトを修正することもできますが、相当な知識と忍耐強い作業を要求されます
- シソーラス辞書を使えばこれらの検索モレをすぐに解決できます

シソーラス辞書を導入して、検索モレのないサイトにしませんか！

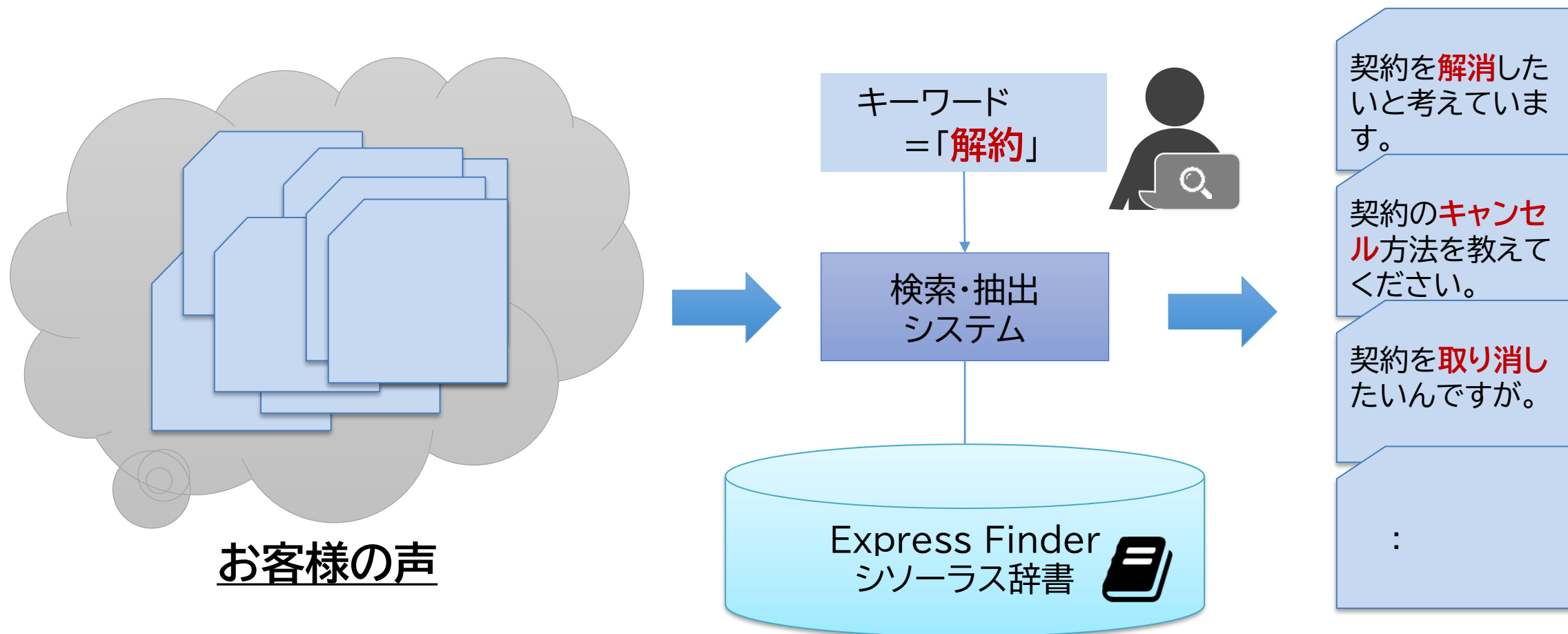
利用シーン2 | シソーラス辞書を活用したスキルマッチング



- 自由な表現で記述される職務記述や職務経歴中の用語のマッチングが実現できます。

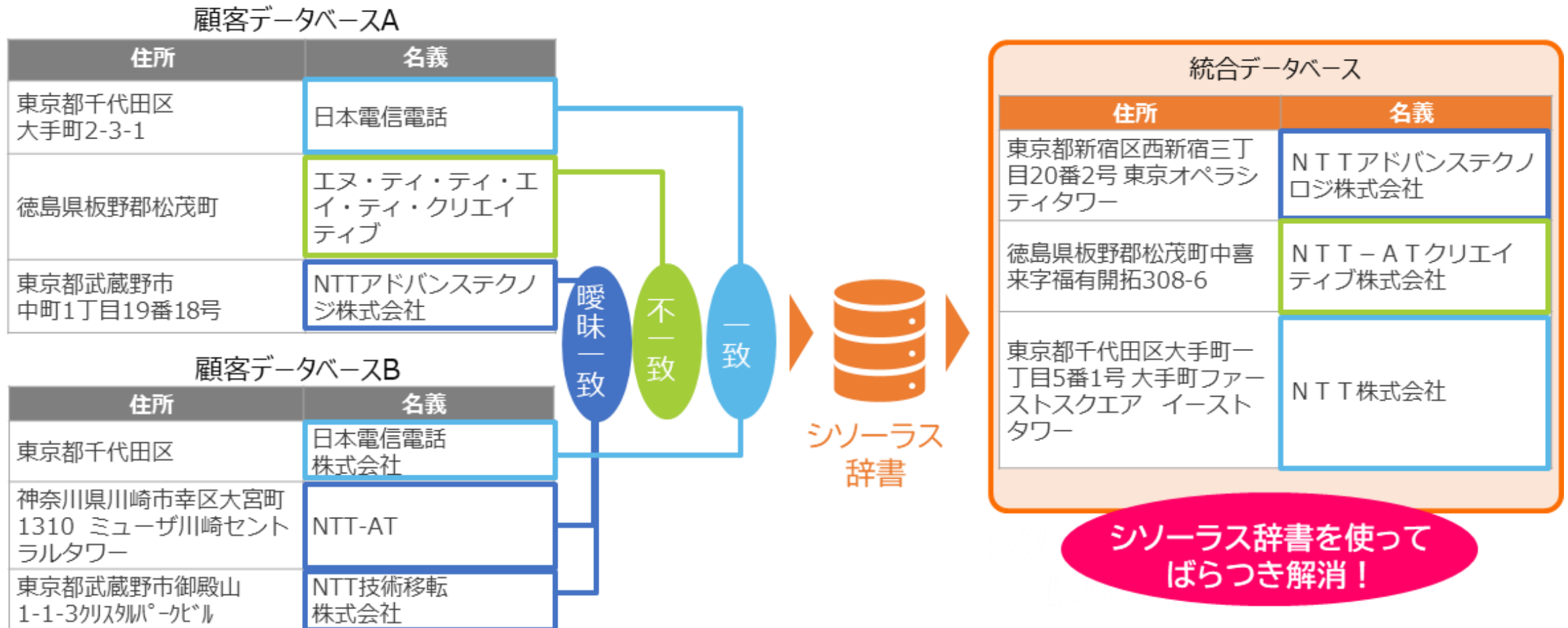


- 多様な表記で記述される「お客様の声」を、幅広く検索・抽出できます。



利用シーン4 | シソーラス辞書を活用した顧客DBの名寄せ

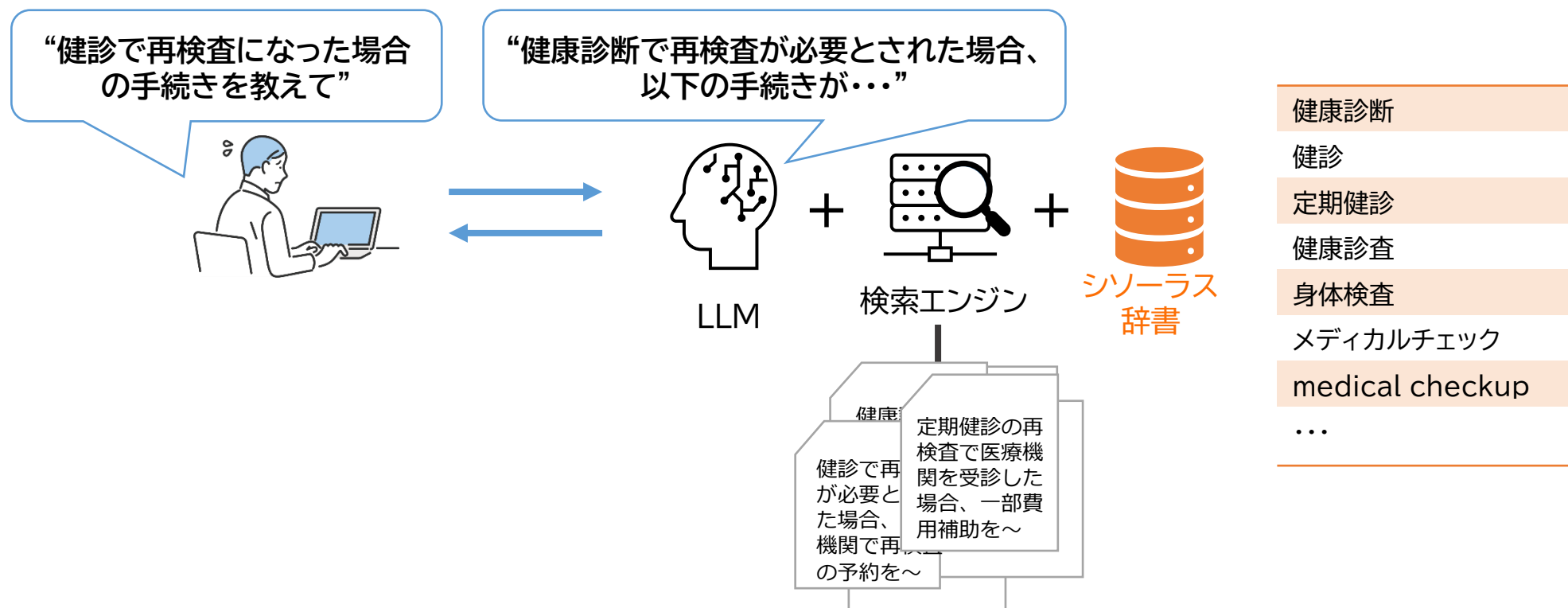
- 旧名、略称、愛称、通称、前株・後株など様々な表記で登録された顧客名を名寄せできます。



利用シーン5 | 大規模言語モデル(LLM)と検索エンジンを融合するRAGシステムの改善

■ 検索漏れを防止し、LLMの回答を改善できます。

- 背景と課題検索エンジンを用いて大規模言語モデル(LLM)に独自知識を追加するRAG(Retrieval-Augmented Generation)技術が注目されています。独自知識の検索には埋込ベクトル検索を用いる場合もありますが、手堅く全文検索エンジンを用いる場合も多くあります。
- 全文検索を用いてRAGシステムを構築する場合、表記ゆれによる検索漏れが課題となります。
- 適切な文書が検索されないとLLMも適切に回答できません。



Express Finder/分冊構成

- Express Finderは分野ごとの分冊構成をとっています。- 用途に併せて分冊を選択いただくことで、意図せぬ同義語展開を避けることができます。
- 詳細な収録例は、HPよりダウンロード可能なサンプルリストをご覧ください。

分野	分冊	語数 (*1)	収録内容	収録例
基本語	基本語	221,524	一般用語を収録。(語義レベルの同義語辞書)	コンピュータ ⇔ computer、電子計算機、電腦
	ことばの表記ちがい	199,251	漢字ひらがなカタカナの揺れや送り仮名の揺れなどの表記ゆれを収録。(表記レベルの同義語辞書)	引換券 ⇔ 引き換え券、引替え券、ヒキカエケン、ひきかえけん
組織名称	企業名	924,582	国内の企業名称を収録。登記名変更や企業合併に伴う名称変更や略称などを収録。	NTTアドバンステクノロジー ⇔ 日本通信技術株式会社 ⇔ NTT-AT
	機関・団体・学校名	299,839	国内の行政機関、団体、学校等の名称を収録。	国立国会図書館 ⇔ NDL
専門用語	経済・法令・公共用語	192,010	経済・ビジネス用語や公文書・法律・行政用語などを収録。	全世界衛星測位システム ⇔ GPS、全世界衛星測位システム
	時事・スポーツ・芸能・放送用語	81,376	時事(ニュース)、スポーツ一般、芸能に関する用語を収録。	タイムテーブル ⇔ ラジオ・テレビ欄、ラテ欄、番組時刻表 ボッチャ ⇔ Boccia、Bocce、ボッチェ
	医学・薬学用語	626,797	医学薬学に関する用語を収録。病名、薬物名、人体の部分名など。	フレイル ⇔ Frailty、虚弱
	非鉄・金属用語	20,683	非鉄金属、金属加工などの用語を収録。	ウッドメタル ⇔ Wood's metal、ウッド合金、易融合金
	コンピュータ・情報・通信用語	131,320	コンピュータ、情報通信、電子工学に関する用語を収録。	液晶ディスプレイ ⇔ リキッド・クリスタル・ディスプレイ、LCD
	自動車用語	51,075	自動車を中心に電車・船などを含めた運輸交通に関する用語を収録。	最小回転半径 ⇔ minimum turning radius、内側車輪最小旋回半径
	機械用語	170,451	機械工学に関する用語を収録。	遠心ファン ⇔ centrifugal blower、輻流送風機
	土木建築用語	83,691	土木、建築に関する用語を収録。岩石、地形、工法、建造物に関連する語句を対象とする。	昇降機 ⇔ elevator、リフト、エレベーター
	電力・エネルギー用語	39,346	環境、発電、原子力、物理学に関連する用語を収録。	地熱ヒートポンプ ⇔ geothermal heat pump、GSHP、地中熱ヒートポンプ
	化学・石油工業用語	183,005	化学分野、主に石油化学工業に関連する語句を収録。	ポリアミド ⇔ polyamide、PA

辞書のご提供形態など

- 辞書はテキストファイル(tsv形式)でご提供。
- 定期的な辞書更新(企業名辞書は年4回、その他分冊は年1回)。
- ご利用方法に合わせてお見積りいたします。

1 ご提供形態

テキスト
ファイル
(tsv形式)

- ・ お客様システムへインポート可能
- ・ システム導入に必要な範囲で改変可能

2 辞書の更新

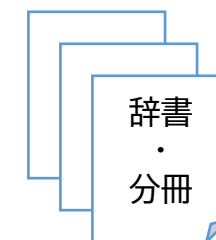
企業名辞書
年**4**回

その他分冊
年**1**回

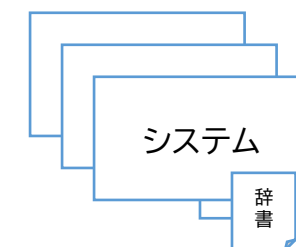
- ・ 辞書は随時更新

3 お見積

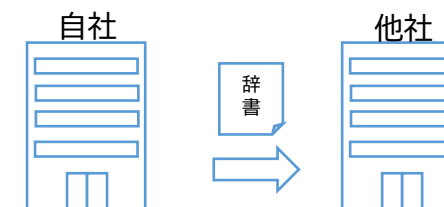
ご利用辞書数



ご利用システム数



自社利用/第三者提供



- ・ ご利用方法に合わせてお見積り
- ・ 契約は原則年単位

お客様と共に進化し続けるバリューパートナー

お客様に真摯に向き合い、時流を読みながら常に革新的なご提案を行い、
新しい価値を生み出し続けるパートナーでありたい、それが私たちの目指す姿です。

