

# テキストマイニング関連製品と ソリューションのご紹介

～テキストマイニング・特許分析・類似検索ソリューション～

数理システムユーザーコンファレンス2014

株式会社NTTデータ数理システム [tmstudio-info@msi.co.jp](mailto:tmstudio-info@msi.co.jp)

1. **TMS**
2. **PMX**
3. **TMX**
4. 類似検索ソリューション



# Text Mining Studio



## TextCutter

ホテル客室  
の話題

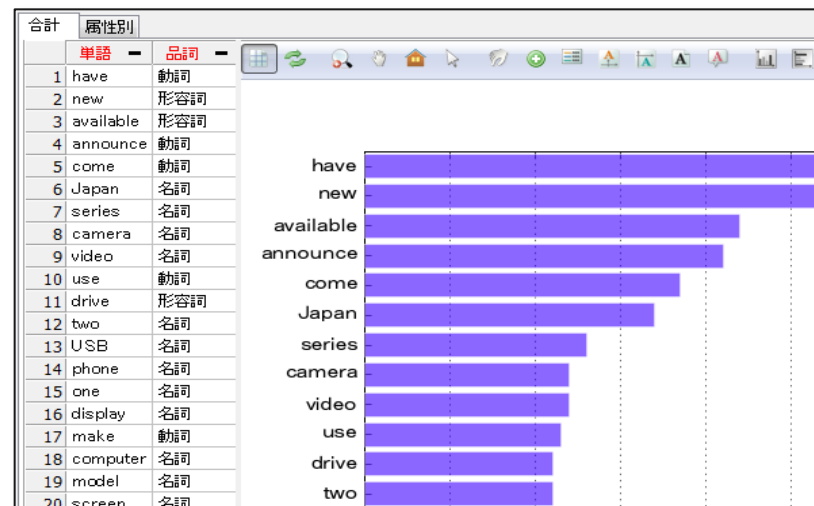
お部屋の方は古臭く、若干狭  
いかな……と感じてしまいま  
した。

付属施設の  
話題

水族館は綺麗  
でデートに最適、イルカショ  
ーも見ごたえあります！

- 雑多なテキストを  
トピック毎に**分割**するツール
- **分析精度向上**のための  
前処理に威力を発揮
- 当社完全オリジナルの技術、  
**類似製品はありません！**

## 英語アドオン



- **TMS** 上で **英文** の  
テキストマイニングが可能に
- **係り受け**の抽出、  
**否定表現**の抽出 にも対応
- 日本語分析と全く同様の使用感

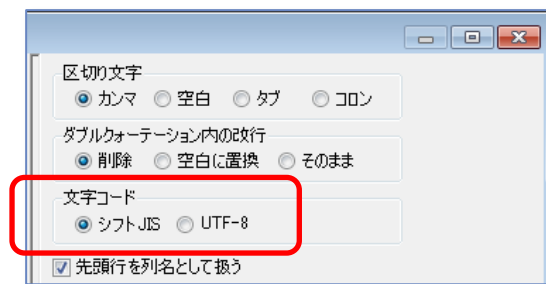
- 入力データ 文字コードの選択を追加
  - **UTF-8 のテキストファイルを取込み可能に**

- 日本語解析 3.2エンジン  
かな-漢字変換の精度向上

- **より一般的、馴染みのある表記に統一するよう修正**

- 64Bit版

- **TMS本体で扱える メモリ空間 が拡大、  
多数の分析結果を作成しても安定して動作が可能**  
※一部、画面インターフェースの使い勝手が異なる部分がございます。





# Patent Mining eXpress

## 簡便な操作で特許文書のマイニングを 特許情報分析ツール

- **WEBブラウザ**上の簡単操作で  
書誌情報・テキスト情報の  
分析・可視化
  - 複雑なパラメタ設定は徹底的に排除
- **TMS**特許情報分析の  
ノウハウを結集
  - 特許文書に依存した **キーワード抽出・可視化のチューニング**を施し、**組み込み**



Patent Mining eXpress (PMX)



## 分析観点

この分野は  
盛り上がっているのか？  
衰退しているのか？  
その中での自社の位置は？

どんな技術が  
どんな課題を解決する  
ために用いられている？

どこがどんな技術を  
持っている？

類似の特許には  
どんなものがある？

出願状況  
時系列把握

発明の方向性  
技術用語抽出

発明マップ

## 意思決定

研究の継続判断  
投入リソース検討

研究の方向性検討

新規応用分野開拓

顧客・共同研究先の  
開拓

# これまでの Patent Mining eXpress

Version 1.0

2013年8月

処理の高速化

Version 1.1

2013年11月

原文参照の強化

Version 1.2

2014年5月

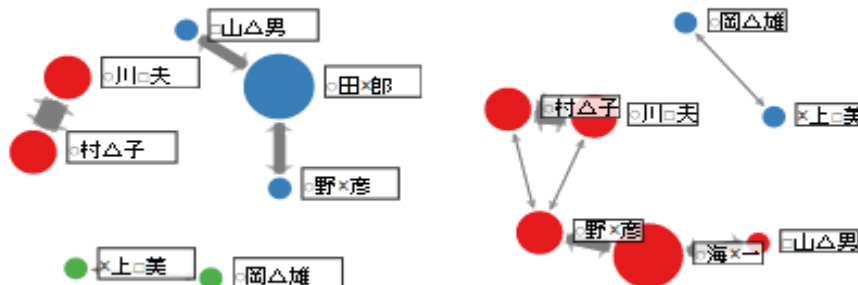
Windows 8.1対応

発明者ネットワークグラフ

Version 1.3

2015年初旬

発明者間の関係が一目瞭然



## 分析データ：データの管理を行う

- ・登録
- ・選択・削除
- ・辞書登録 ◎次VU予定

分析対象データを登録  
分析対象データの変更・削除  
技術単語の削除・まとめ上げの指定

## 出願状況：書誌情報に着目し、出願の状況を図示する

- ・出願の動向を知る
- ・出願人に着目する
- ・発明者に着目する
- ・分類コードに着目する

出願人数や発明者数、出願件数などの推移を図示  
出願人内訳の推移を図示  
発明者内訳の推移を図示  
分類コード内訳の推移を図示

## 発明の方向性：特許のテキスト情報からクロスマップを作成する

- ・課題と解決手段の関係を知る
- ・出願人と技術の関係を知る
- ・その分野の技術動向を探る ◎次VU予定

課題と解決手段のキーワード間の関連を図示  
出願人とキーワードとの関連を図示  
分類コードと解決手段との関連を図示

## 発明マップ：発明内容をマッピングする

- ・特徴技術ネットワーク
- ・ポジショニングマップ

出願人と特徴技術間の関連をネットワークにより図示  
特許テキスト間の類似度距離マップを図示



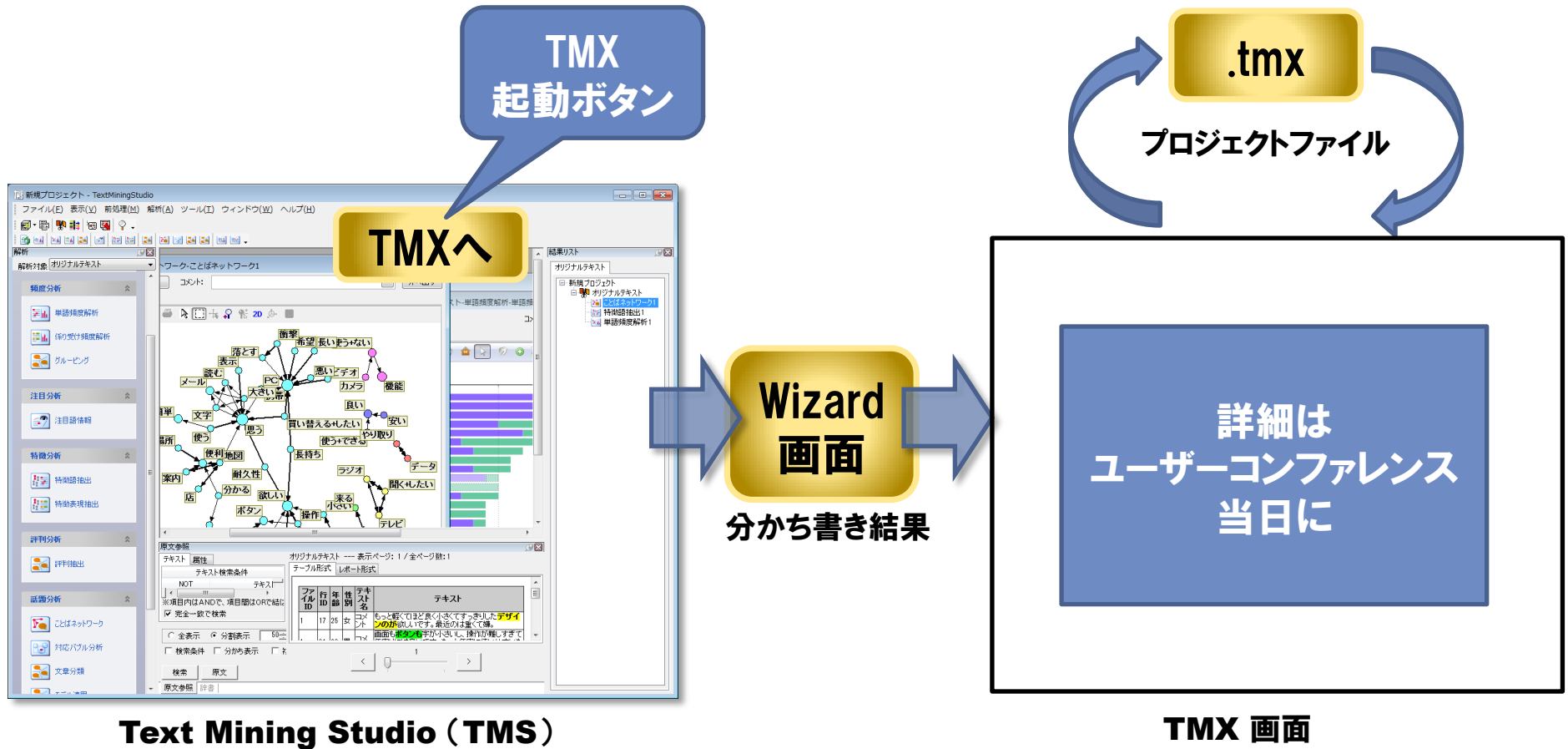
# Text Mining eXpress

## テキストマイニングをよりお手軽に

以下のコンセプトの元、開発しています！

- 簡単操作
  - まず結果を表示。結果を見ながらの条件変更
  - 当社BigDataModuleなどの高速結果作成技術の適用
  - クリック、ドラッグ&ドロップ操作の活用
- 分析結果から分析結果へ、よりスムーズに
  - 次の分析への「気づき」のためのヒント画面
  - 分析条件の引き継ぎ
  - TMSとは別の視点の分析のご提供

## ファーストリリース: TMSから起動します!





# 類似検索ソリューション

## 当社のテキストマイニング技術を用いた 類似検索ソリューション

### ・自然文による検索

- 類似度の高い順に該当文書を複数件数表示
- コールセンター、FAQシステムでのご活用
- 単語による検索も可能

通常の検索の場合の例：

故障 返金

検索



FAQ検索結果

Q:故障、返金について教えてください  
A: 次の窓口へ問合せください。  
03-0000-0000

該当項目が少なすぎる、電話をかけても繋がらない、または「検索結果は 0件 です。」と出てしまう。

数理システム類似検索ソリューションを使うと・・・



故障 返金

検索



故障したので返金したい

検索

Q:故障、返金について教えてください  
A: 次の窓口へ問合せください。  
03-0000-0000

Q:型番A□の取扱代理店はどこにありますか

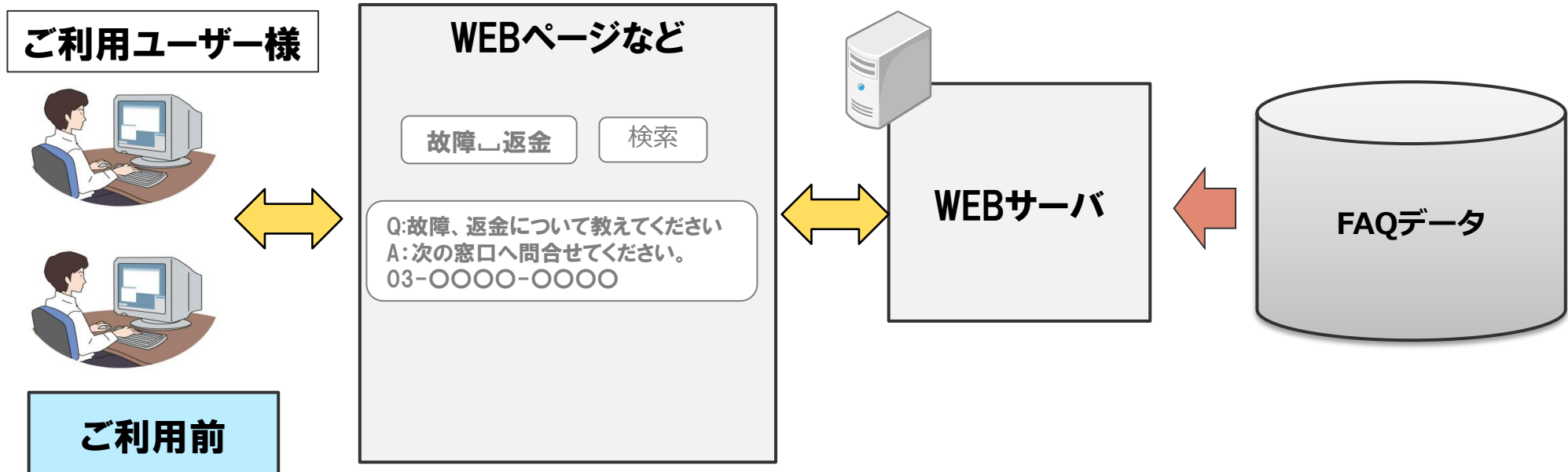
A: 代理店一覧はこちらです  
<http://>

Q: . . .

A: . . .

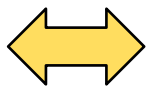
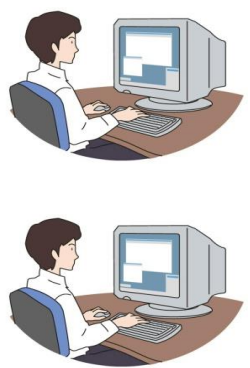


# 類似検索ソリューション～ご利用イメージ



# 類似検索ソリューション～ご利用イメージ（2）

ご利用ユーザー様



ご利用後

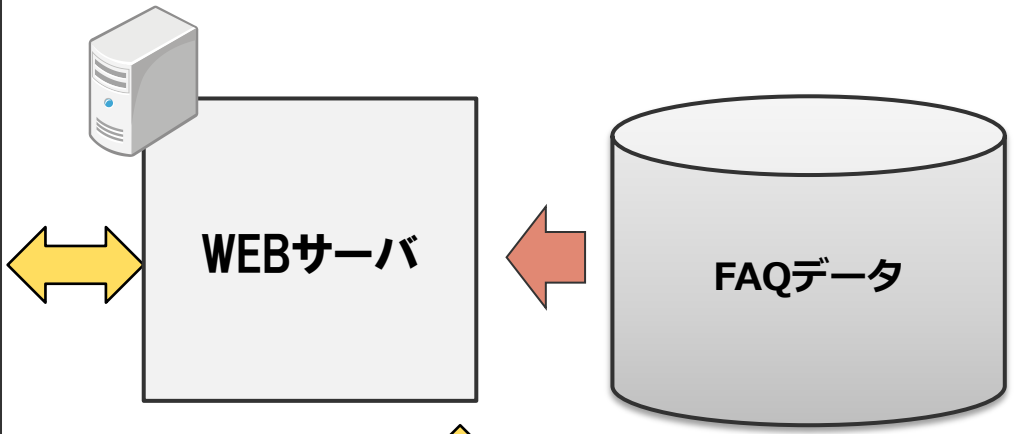
**WEBページなど**

故障したので返金  
したい

Q:故障、返金について教えてください  
A: 次の窓口へ問合せください。  
03-0000-0000

Q:型番A□の取扱代理店はどこに  
ありますか  
A: 代理店一覧はこちらです  
<http://>

Q:・・・  
A:・・・



類似検索  
情報授受

分析担当者様



**FAQ分析サーバ**

**TMS**

ユーザー  
辞書

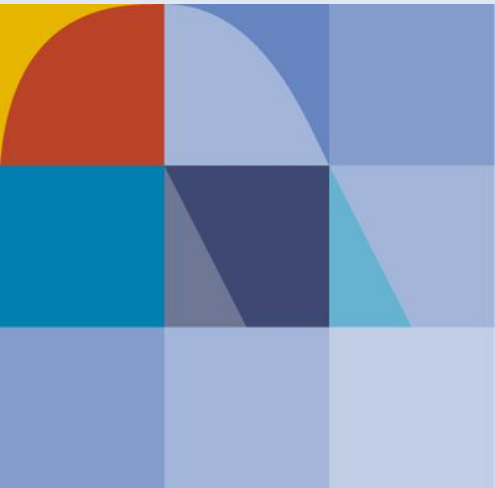
**TMSプロジェクト  
辞書情報**

**類似検索  
情報作成**



# 類似検索ソリューション～主な仕様

- 入力：
  - 自然文、キーワード列
- 出力：
  - 類似度付き文書データ
    - お客様の環境に応じカスタマイズ致します。
- 類似度計算時処理：
  - TMSを利用した類義語辞書の適用(類義語自動抽出)
  - 漢字、かな、カナの吸収(オプション機能)
  - その他オプションも開発中
- 環境：
  - Windows
  - データベース(ベンダーはご相談)
  - TMS



Global IT Innovator  

---

NTT DATA Group **NTT DATA**