

テキストマイニングとその応用技術 ～新製品・特許分析ツールPMXのご紹介～

数理システムユーザーコンファレンス2013
株式会社NTTデータ数理システム
tmstudio-info@msi.co.jp

1. **Text Mining Studio (TMS)**

- **TMSとは**
- **アドオンモジュール**

2. **Patent Mining eXpress (PMX)** 概要

- **PMXとは**
- **特許分析のねらい**
- **PMXの基盤技術**

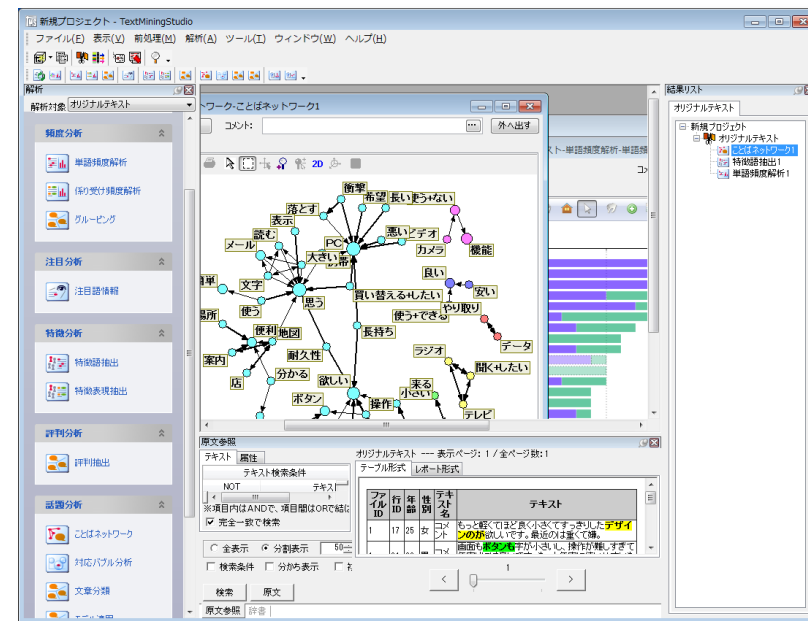
3. **PMX** を使う

- **機能紹介**

4. リリーススケジュール・今後の展望

テキストデータから有益な情報を抽出するための テキストマイニングツール

- **誰にでも**高度なテキストマイニングを
 - マウスの操作のみで分析が可能
 - 豊富な分析機能 と 強力なグラフ機能
- より**自由度**の高い分析を
 - カテゴリ機能等「意味」に着目した分析
 - データマイニングツール **Visual Mining Studio** とのシームレスな連係



Text Mining Studio (TMS)

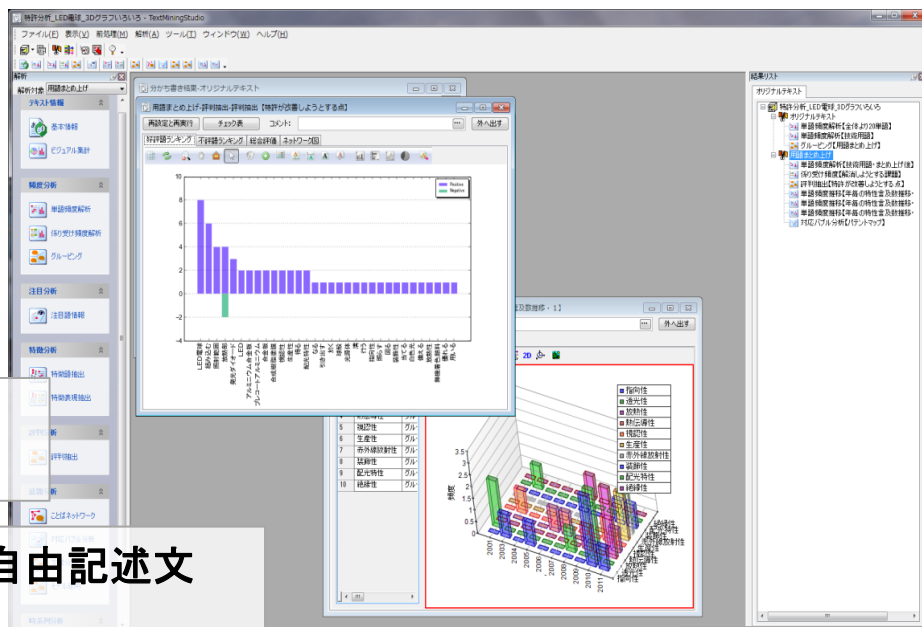
これ1つでマルチシーンに対応する
テキストマイニングツールです！

コールセンター
・マーケティング
・離反回避

営業・業務日報
・社員育成

アンケート自由記述文
・商品開発
・製品弱点の発見

SNS分析
・評判分析
・他社動向把握
・広告効果把握
データ販売ご相談ください



医療・看護系
・医療サービス向上
・医療危険予測

特許明細書
・技術マーケティング

新聞・雑誌記事
・社会動向の把握
・海外マーケティング

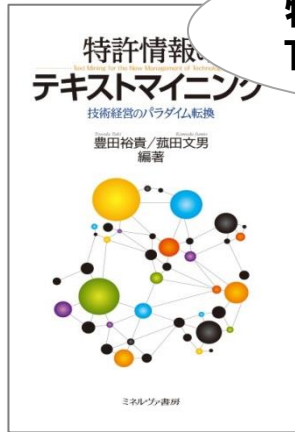
学術系

～詳細はパンフレットをご参照ください～



各社のテキストマイニングシステムの紹介と比較、さらにテキストマイニングやデータマイニングに役立つ情報を掲載。

ISBN 978-4-320-12200-0



特許分析
TMSチュートリアル付

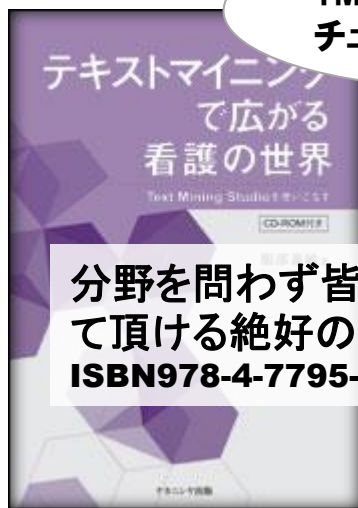
花王・P&G・ユニリーバ
の4年分の特許データから
各社を比較、分析。
ISBN: 9784623059768



マーケティング
SNS分析

人気項目：
「冬に日焼け止めは売れる
のか」

ISBN: 4-495-64281-5



TMSトライアル版/
チュートリアル付き

分野を問わず皆様に使っ
て頂ける絶好の入門書！
ISBN978-4-7795-0511-9



看護研究

第38章 Text Mining Studio
による言語の量的分析

ISBN978-4-86174-035-0



社会研究

18章 テキストマイニング (著者 稲葉光行)
1. テキストマイニングの目的 / 2. データ収集 /
3. データ分析法 / 4. テキストマイニングの利点
と限界
テキストマイニングの研究例 テキストマイ
ニングによる非営利ネットワークの組織化プロセス
の可視化
ISBN978-4-7795-0546-1

2013年業界初 ソフトウェア賞 受賞！

**TMSは、機能の豊富さ、使いやすさ、他製品とのシームレス連携機能と
いった実用的価値を評価され、日本計算機統計学会の「2013年度 日本
計算機統計 学会 ソフトウェア賞(開発賞)」を受賞しました。**

なお、TMSのエンジンとして稼働している同社のデータマイニングツール、
Visual Mining Studio(通称VMS)も2004年に同賞を受賞しており分析業
界において数理システムの強さが改めて証明された事となります。

～TMSユーザーの皆様へ～

日頃ユーザーの皆様からはTMSの機能において、貴重な
ご意見を頂き、誠にありがとうございます。今回の受賞を
機に今後もますます皆様からのご意見をバージョンアップ
に反映し、引き続き業界最多機能、そして使いやすさにこ
だわって開発をすすめて参ります。

2013年2月、使いやすさを追及し生まれ変わったTMSを、
今後とも末永くご愛用頂きますよう、よろしくお願い致し
ます。

TMS開発チーム一同

表彰状

ソフトウェア賞(開発部門)

(株) 数理システム

Text Mining Studio

貴社が開発された上記ソフトウェアは、テキスト
マイニングに関する多様な機能、使いやすいユー
ザインタフェース、他製品との連携機能などを合
わせもつ優れた製品であることが、本学会での展
示や講演などを通じて広く認められます
ここに賞を贈り敬意を表します

2013年5月16日

日本計算機統計学会

会長 石橋 雄

TextCutter

ホテル客室
の話題

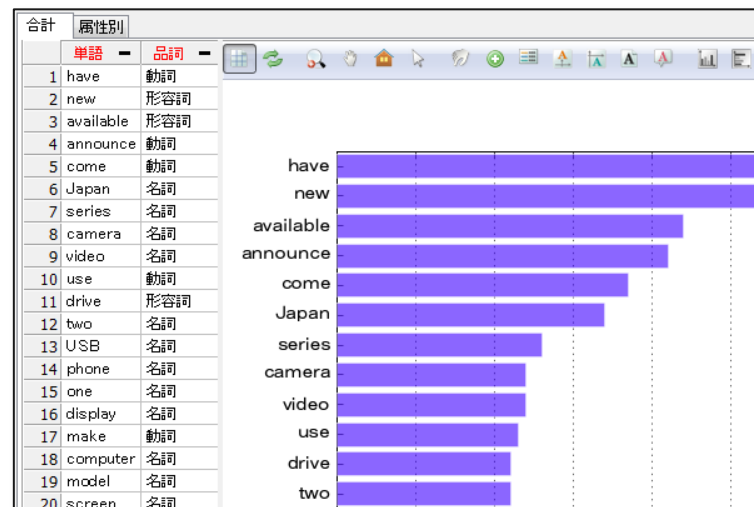
お部屋の方は古臭く、若干狭いかな・・・と感じてしまいました。

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

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)

公開特許公報

特許公開出願番号

特開2013-XXXXXX

公開日 平成25年11月22日

Int.Cl.

FI

M01A 01/01

M01A 01/01

M02A 01/01

M02A 01/01

出願人 ●●食品工業株式会社

発明者 ○田×郎

発明者 ○野×彦

【発明の名称】 食感を高めた豆腐の製造方法

【要約】

【課題】

豆腐の食感を高めることを目的とする。

【解決手段】

製造過程で天然抽出物Aを加えることを特徴とする。

.....

特許公報情報の例（大幅に簡略化）

出願番号

- 1件の特許を特定するキーとなる

公開日

- 時系列推移を測るキーとなる

分類コード

- 技術要素と密接な関係にある

出願人・発明者

- 技術動向把握の上で重要な要素
- 一般に複数の発明者・出願人が1件の特許に対応している

テキスト情報

- 【要約】部分は、発明の概要・要点が簡潔に記されており、内容把握のうえで重要な情報
- 具体的内容は、続く【請求の範囲】を参照する

書誌情報

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

Patent Mining eXpress 機能一覧

分析データ：データの管理を行う

- ・登録
- ・選択・削除

分析対象データを登録
分析対象データの変更・削除

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

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

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

かゆい所に手が届くツールを

とにかく簡単であること

自由度はやはり必要だ

マルチユース

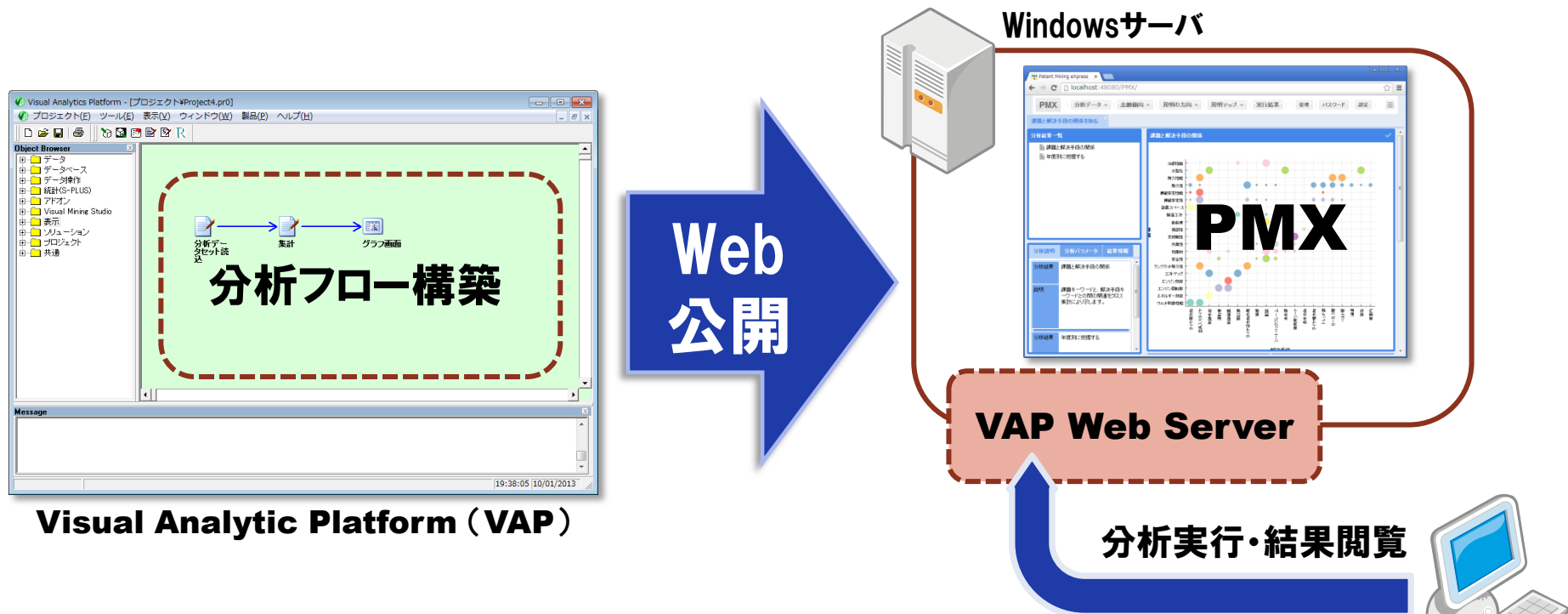
有用な結果が
ダイレクトに欲しい

アルゴリズムまで知りたい

目的別・専用ツール

データマイニングとも連携させたい

PMX の実現基盤：プラットフォームVAP



分析プラットフォーム製品 **Visual Analytic Platform**

- 独自の**分析フロー**を構築し、これを**Webアプリケーション**として公開することが可能

- 特許分析に特化した**独自のロジック**を開発し、
Webアプリケーションとしてパッケージ化したものが**PMX**

PMXを使う ～データの取込み～

PMX

分析データ

出願動向

発明の方向

発明マップ

登録

選択 削除

admi さん

新たに分析を実行済み

管理者用の

分析パラメータ

ファイル名: C:\fakepath\sample.csv

ファイルを選択

CSV読み込み設定

データプレビュー

文字エンコード: Shift_JIS

列番号	元列名	列名
1	出願番号	出願番号
2	出願日	出願日
3	発明の名称	発明の名称
4	出願人	出願人
5	発明者	発明者
6	Fコード	Fコード
7	課題	課題
8	解決手段	解決手段

実行

Copyright © 2013

A	B	C	D	E	F	G	H	
1	出願番号	出願日	発明の名称	出願人	発明者	Fコード	課題	解決手段
1	特許出願2012-W4mcaC	20121221	大豆の製造方法	●●食品工業株式会社	○田×郎 ○野×郎	M01A 01/01 M02A 01/01 M03A 01/01	大豆の食感を高めることを目的とする。	製造過程で天然抽出物Aを配合することを特徴とする。
2	特許出願2012-vAhlzC	20120916	大豆の製造方法	株式会社○●製菓所	○村△子 ○山▽夫	N01A 01/01 N02A 01/01	大豆の製造コストを削減する。	製造方法及び製造方法を適用することでコストを抑える。
3	特許出願2011-gHjzC	20111008	長期保存可能な大豆	株式会社●●製菓所	×上口美	M01A 01/01 N02A 01/01	大豆の長期保存性を高めること。	製造過程で殺菌技術を適用することによって特徴とする。
4	特許出願2011-vAhlzC	20110817	大豆の製造方法	●●食品工業株式会社	○田×郎 ○山口△夫	M01A 01/01 M02A 01/01 M03A 01/01	食感を高め、歯ごたえを改善した大豆を提供する。	天然抽出物Aおよび天然抽出物Bを配合することにより、良好な食感を得る。
5	特許出願2010-vAhlzC	20110622	大豆の製造方法	株式会社○●製菓所	○村△子 ○山▽夫	N01A 01/01 N02A 01/01	製造コストを抑えた大豆の製造方法。	製造方法で製造コストを抑えることの特徴とする。
6	特許出願2010-kRrc94	20100419	長期保存可能な大豆	株式会社●●製菓所	○田×郎 ○野×郎	M02A 01/01 M03A 01/01	大豆の食感向上させること。	天然抽出物Bを添加することを特徴とする。
7	特許出願2010-gHjzC	20100307	長期保存可能な大豆	株式会社●●製菓所	×上口美 ○山△雄	M01A 01/01 M02A 01/01	大豆の長期保存性を高めること。	殺菌技術の適用することで、長期保存が可能とする。
8	特許出願2010-RgY57v	20100129	大豆の製造方法	●●食品工業株式会社	○田×郎 ○山口△夫	M01A 01/01 M02A 01/01 M03A 01/01	大豆のより良い歯ごたえを実現すること。	天然抽出物Aおよび天然抽出物Bを配合することで、良好な歯ごたえを得る。

表形式に成形された
特許情報データ

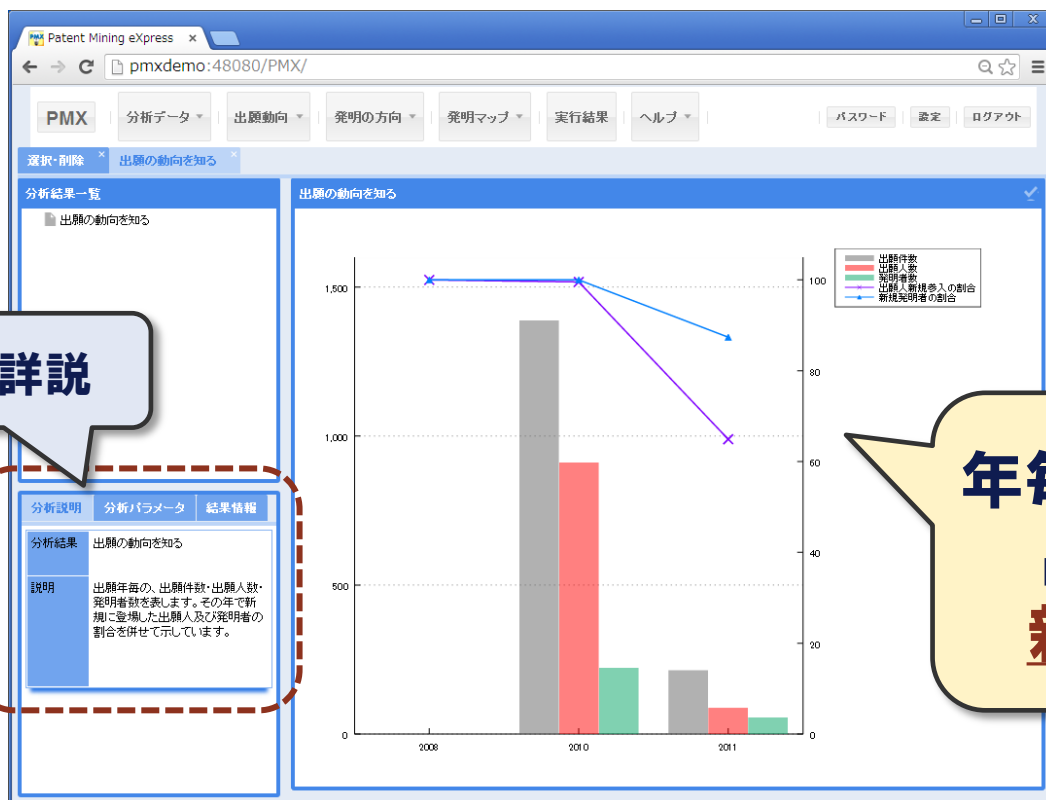
- お手元の .csv ファイルをアップロード
 - 列名から 書誌情報・テキスト情報 の列を自動判定し、
テキストの 分かち書き処理 を自動的に実行

PMXを使う ～出願動向・1～

出願動向

- 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

- メニュー「出願動向」から
全体傾向・出願人・発明者・分類コード
の各種 属性 に着目した傾向把握が可能

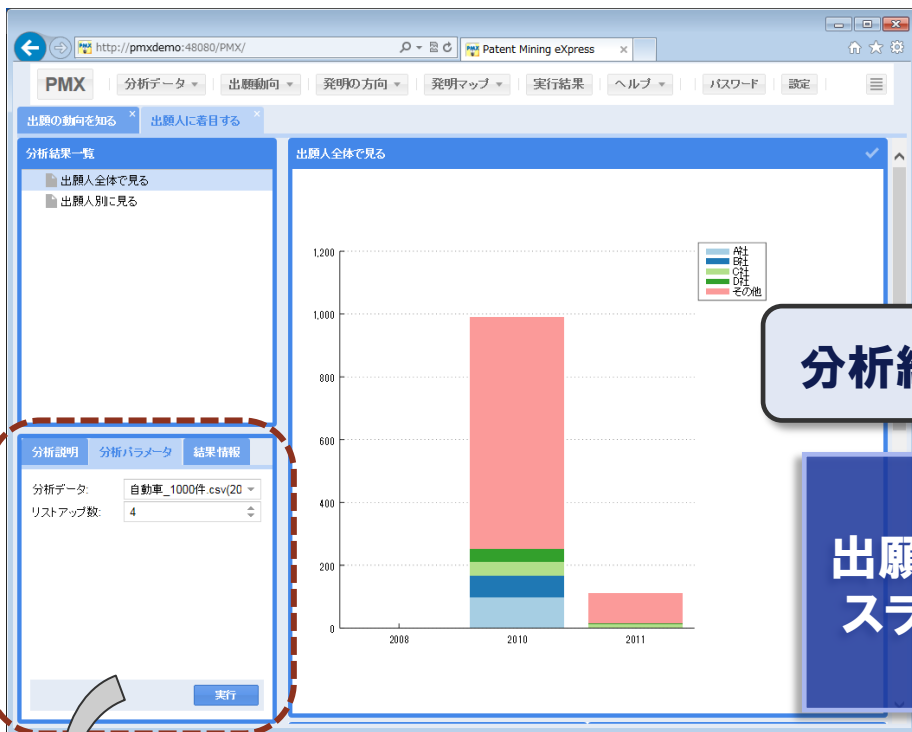


分析の詳説

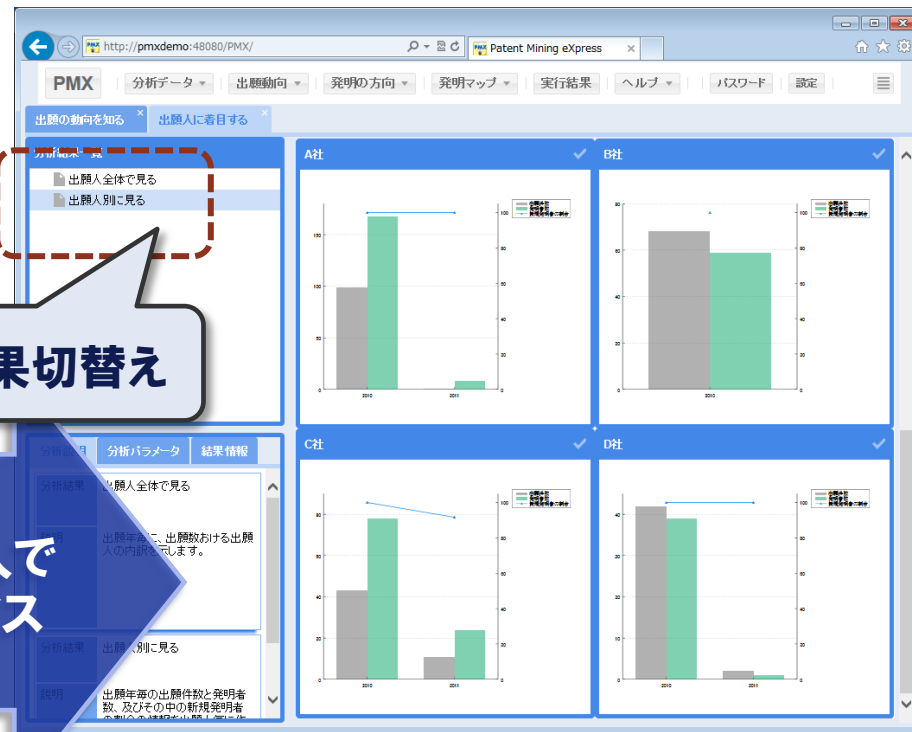
年毎の件数にあわせて、
出願人・発明者の
新規参入数を図示

「出願の動向を知る」(全体傾向)

PMXを使う ～出願動向・2～



「出願人に着目する」全体



「出願人に着目する」個別

分析パラメータ

- 必要であれば 分析パラメータ を変更して再実行することが可能 (通常はその必要はない)

PMXを使う ～発明の方向～

- メニュー「発明の方向」から

課題・手段のキーワードクロス集計 出願人とキーワードのクロス集計

が可能

発明の方向 ▾

課題と解決手段の関係を
知る
課題と出願人の関係を
知る

課題テキスト

【発明の名称】食感を高めた豆腐の製造方法

【要約】

【課題】

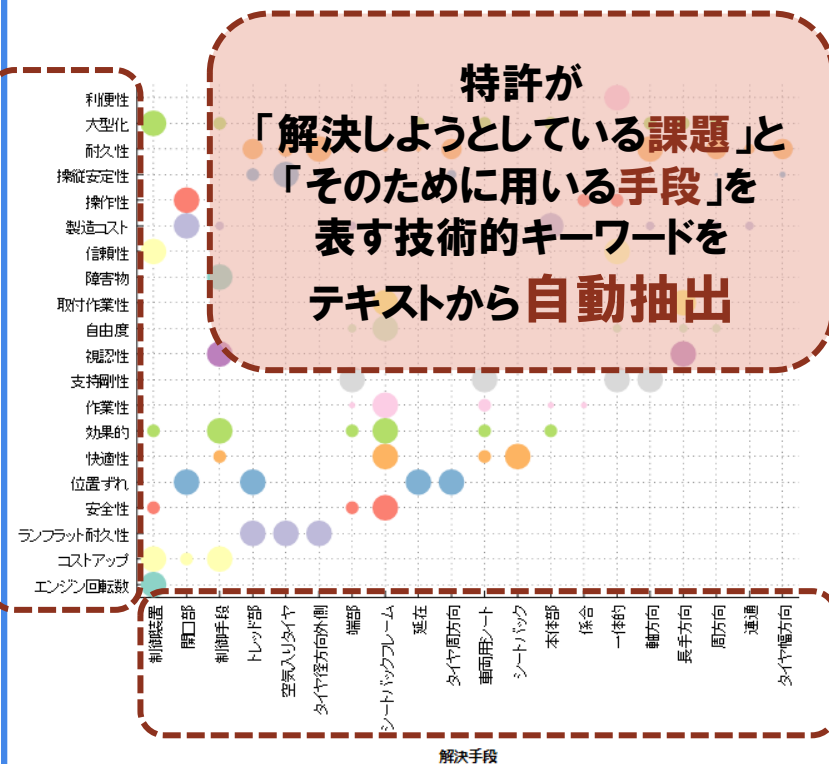
豆腐の食感を高めることを目的とする。

【解決手段】

製造過程で天然抽出物Aを加えることを特徴とする。

手段テキスト

課題と解決手段の関係



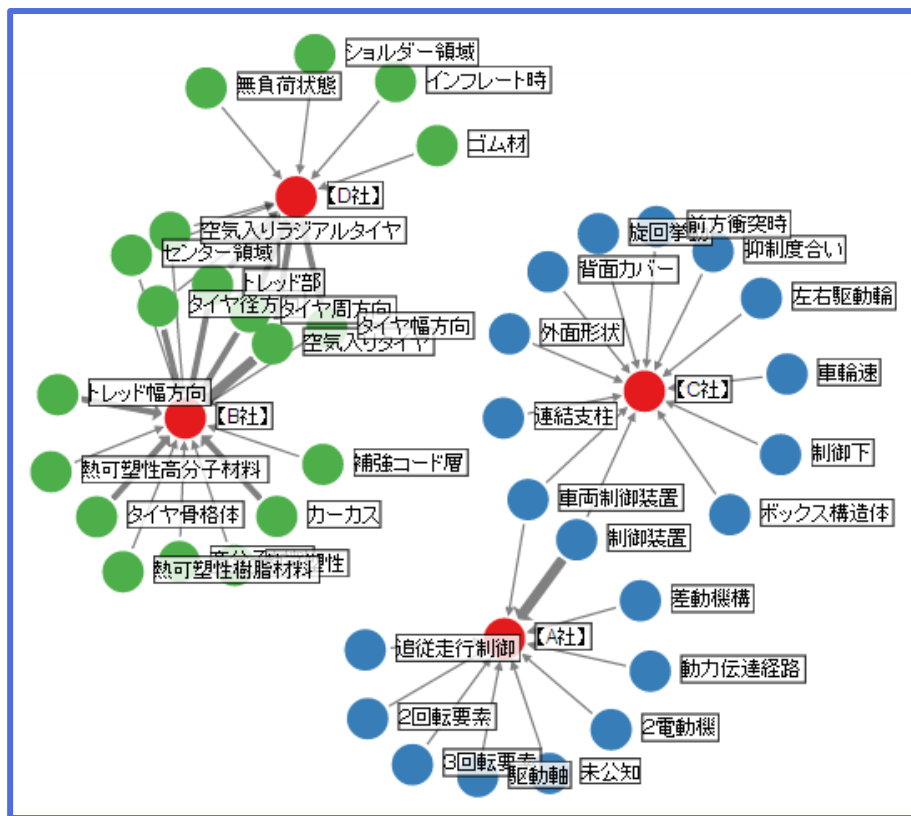
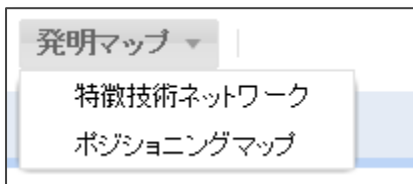
「課題と解決手段の関係を
知る」

PMXを使う ～特徴技術ネットワーク～

- メニュー「発明マップ」から

出願人・技術間のネットワーク図示 特許類似度に基づいた二次元マッピング

が可能

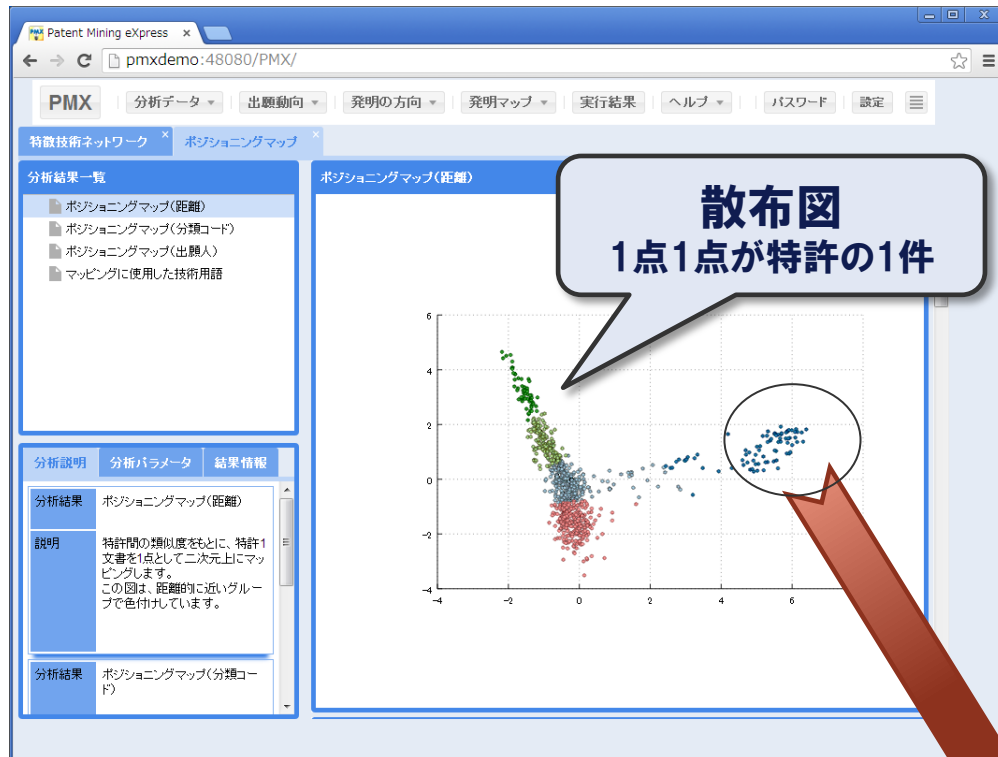


ある出願人において、
こういった技術的キーワードが
特徴的に出現しているか？

技術的キーワードと
出願人との関連を示す
ネットワークグラフにより
これを表現

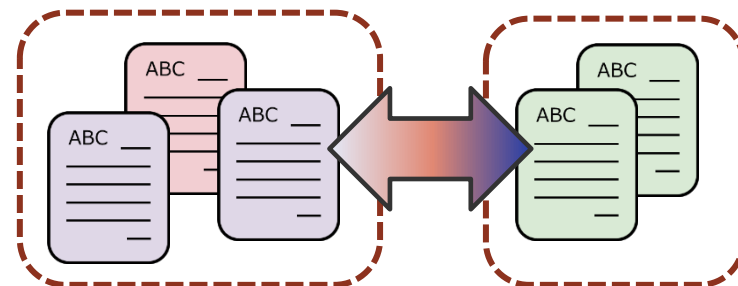
特徴技術ネットワーク

PMXを使う ～ポジショニングマップ～



ポジショニングマップ

- 分析結果のグラフから、**元データ・原文**を確認可能



- 特許文書間に**類似度**を定義し、類似度が高い特許同士がマップ上で近くなるよう2次元平面上の位置を算出する
- 平面上のクラスター、分類コード、出願人**の違いでマップに色付けをして表示

出願	発明	Fコード	公報種別	発明の	審査請求	審査請求	筆頭テ	請求項数	出願年	先頭	
2010070	...	A63...	公開特許公...	遊...	1	false	2C088	1	2010	課	
20110916	...	A63...	公開特許公...	遊...	1	true	2C088	1	2011	課	
20110916	株式...	...	A63...	公開特許公...	遊...	1	true	2C088	1	2011	課
20110922	株式...	...	A63...	公開特許公...	遊...	3	true	2C088	3	2011	課
20110922	株式...	...	A63...	公開特許公...	遊...	3	true	2C088	3	2011	課
20100713	株式...	...	A63...	公開特許公...	遊...	1	false	2C088	1	2010	課
20110926	株式...	...	A63...	公開特許公...	遊...	3	true	2C333	3	2011	課
20100720	株式...	...	A63...	公開特許公...	遊...	1	false	2C088	1	2010	課
20100720	株式...	...	A63...	公開特許公...	遊...	1	false	2C088	1	2010	課
20100721	株式...	...	A63...	公開特許公...	遊...	1	false	2C088	1	2010	課
20100721	株式...	...	A63...	公開特許公...	遊...	1	false	2C088	1	2010	課
20100721	株式...	...	A63...	公開特許公...	遊...	1	false	2C088	1	2010	課

原文参照画面

Text Mining Studio


- ・ **V4.2** に対する修正版 **V4.2.1** 更新ツール リリース
- ・ 2014年リリース 次期バージョン **V5.0**
 - ・ 分かち書き並列実行による高速化、64Bit化 を予定

Patent Mining eXpress

- ・ 2013年9月 **V1.0** リリース ~ 11月修正版リリース

- **PMX**の**ユーザビリティ** 追及路線を踏襲、
複雑なパラメタ設定を排除し
有益な結果に**ダイレクト**に迫れるような
テキスト分析ツールの開発を推進
- 大規模なテキストデータに対する、
ソリューションの**頑健性**・**安定性**の
さらなる向上

JDream III

Jdream IIIとは:2013年3月末まで(独)科学技術振興機構(JST)が管理運営していた日本最大の科学技術文献データベースシステムJdream II。  JDream II
2013年4月からJdream IIIとして(株)ジー・サーチが運営開始。

数理システムでは、弊社パッケージソフトウェア(※)をご利用の方向けに、Jdream IIIデータを分析用データとして販売しております。

(※S-PLUS、Visual Mining Studio、**Text Mining Studio**)

通常のデータベース利用(従量料金制・年間固定料金制)では、紙ベースへの印刷とディスプレイ上への表示による閲覧のみが許可されています。しかし、数理システムより分析用データとしてご購入いただきますと、データの永久保存、データ分析への利用および分析結果の書籍への掲載、論文への発表などが可能です。

☆ **Text Mining Studio**でご利用いただく際には、JDream IIIデータを手軽に分析していただける分析手順の冊子を無料でご提供！

ご購入いただいたデータを入力データとして設定していただくだけで、すぐに一通りの分析が可能です。

1か月間のトライアルを無料でご提供！

テキストマイニングツール for Windows

株式会社 NTTデータ 数理システム

Text Mining Studio



実際に
ご契約頂いた際と同様に
ご質問などには随時サポート！

主な分析機能

- ・テキスト情報
- ・頻度分析
- ・注目分析
- ・特徴分析
- ・評判分析
- ・話題分析
- ・時系列分析

Patent Mining eXpress(PMX)

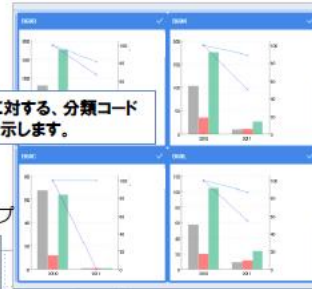
特許情報分析ツール



数理システムText Mining Studioのノウハウが集結！
ブラウザで手軽に利用できる本格派特許マイニングツール
Patent Mining eXpress(PMX)



▼分類コード別

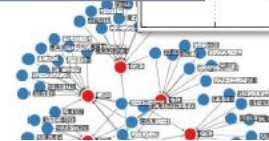


ショニングマップ

出願年毎の出願件数、及びその出願者の割合を、出願年毎に分類コードごとに示します。

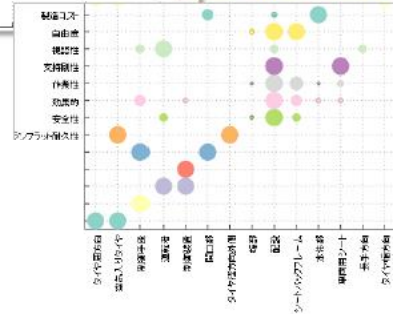
▼特徴技術

出願人毎に特有な技術キーワードがどのように出現しているか、その関係をpネットワーク図にて示します。



▼課題と解決手段の関係

課題キーワードと、解決手段キーワードとの間の関連をクロス集計により示します。



この機会にごゆっくりお試しください

テキスト分析 ご相談ください！



コールセンター



SNS分析



学術系



新聞・雑誌記事

NTT DATA

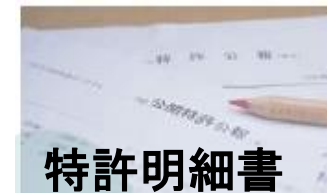
株式会社 NTTデータ 数理システム

TEL : 03 - 3358 - 6681
FAX : 03 - 3358 - 1727

【URL】 <http://www.msi.co.jp/tmstudio/>

【E-mail】 tmstudio-info@msi.co.jp

各種カスタマイズ・
コンサルティングも承ります



特許明細書



医療・看護系



アンケート自由記述文



営業・業務日報