

### Derwent Innovationにおける テキストマイニング機能の効果的な使い方

- 魅せる テーマスケープ マッピング -

2021年4月1日(木)13:30~

クラリベイト・アナリティクス・ジャパン株式会社 ソリューション コンサルタント 宮田 和彦 ソリューション コンサルタント 大谷 美智子



### **Agenda**

- はじめに
- Derwent Innovationでの自然言語処理
  - クラリベイトの製品とサービス
  - テーマスケープ(ThemeScape)
  - テキストクラスタリング
  - 言語解析対象(DWPIの活用)
  - 言語解析処理(しくみ)
- テーマスケープ描画におけるコッやワザ
  - マップ全体像の調整
  - 傾向を際立たせるマーキング
- ストップワードの設定
- カスタムフィールドの活用(ダミー公報)
- おすすめセミナー動画の紹介(オンデマンドセミナー)



### はじめに

### なぜ、テキストマイニングが必要か?

科学技術のイノベーションが加速 量的側面 扱うべき技術情報(特許・文献)が増大 **一情報担当者や研究開発者の負担増** 

質的側面

分野横断的研究の進展 技術やサービスの融合が加速 →**分類化類型整理が追いつかない。** 



必要性を感じるものの→高いハードルを感じる。

- ・どのように活用したら良いかが分からない。
- ・分析や解析の妥当性を検証できない。
- ・どういうキーワードに着目したら良いか分からない。

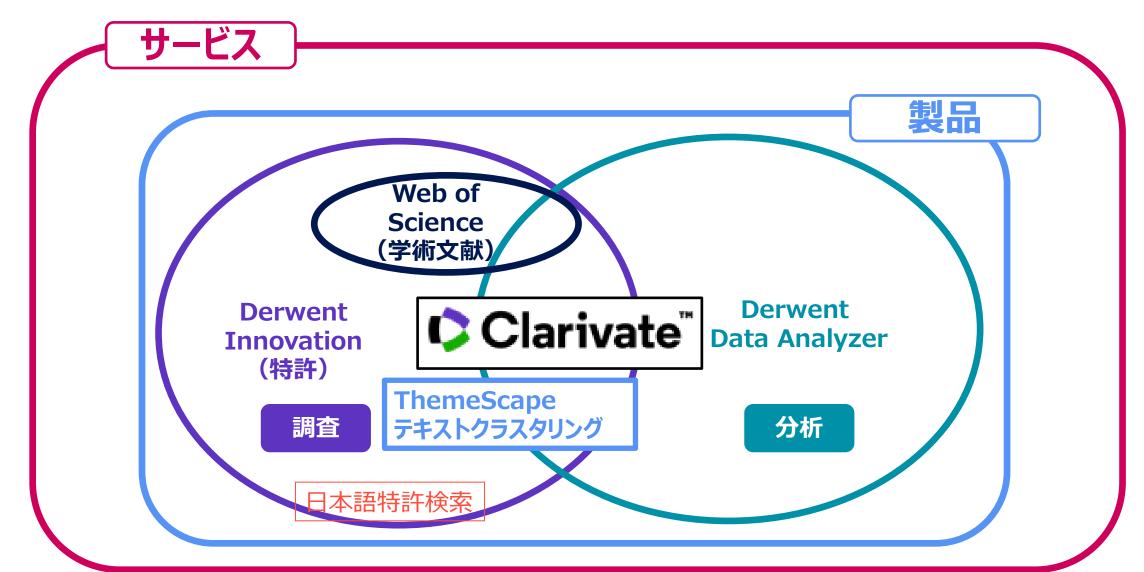


# Derwent Innovationの自然言語処理 テーマスケープとテキストクラスタリング



### クラリベイトが提供する特許関連製品とサービス

クラリベイトでは、特許関連製品として以下のものをご用意しており、これらの製品使った分析サービスもご提供しています。

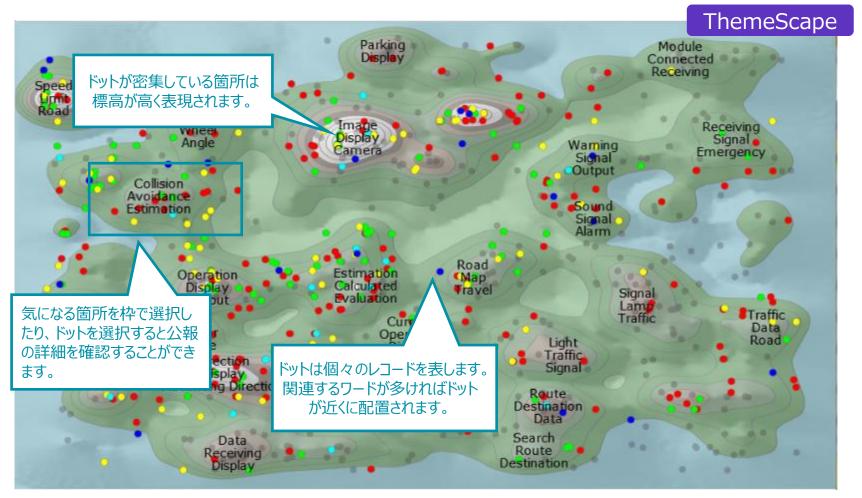




## テーマスケープ(ThemeScape)

ThemeScapeとは、Derwent Innovation 特許データ、DWPI の付加価値特許データ、または科学文献コンテンツから情報をテキストマイニングし、マップを作成するデータ解析ツールです。

解析結果は等高線マップで表現されます。得られたマップに対して、さらに数多くの項目で再検索でき、様々な角度から俯瞰的に知的財産に関する全体像を俯瞰することができます。





<sup>※</sup>テーマスケープは、Analystレベルのご契約がないと利用できません。

### テキストクラスタリング

テキストクラスタリングは、Derwent Innovation 特許データ、DWPI の付加価値特許データ、または科学文献コンテンツから情報をテキストマイニングし、頻出の単語やフレーズ毎にレコードを階層状に分類する機能です。ユーザーが選択したフィールド中のテキストを自然言語解析して、特許や文献レコードを自動的に分類し、結果が階層状のツリー表示されるツールです。解析されたワード情報を見ることで母集団の特徴を把握したり、ワードをクリックすることで仕分けされた特許を抜き出すことが

できます。



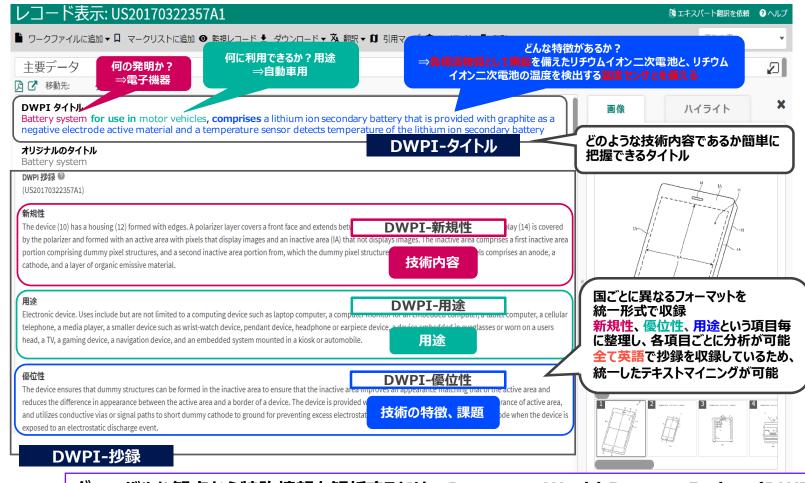


<sup>※</sup>テキストクラスタリングの処理数は、10件以上、1万件以下です。

<sup>※</sup>テキストクラスタリングは、Analystレベルのご契約がないと利用できません。

### **Derwent Innovationの言語解析対象**

Derwent Innovationでは、ThemeScapeやテキストクラスタリングで解析するテキスト対象として**DWPI抄録**を使用することができるので、より目的にフォーカスした解析が可能です。



世界中の特許を統一した言語(英語)で、 一貫性をもった基準で新規性、優位性、 用途と項目ごとに整理された

DWPI抄録を解析することで、 出願国に関わらず、項目ごとの解析が 可能となります。

グローバルな観点から特許情報を解析するには、Derwent World Patents Index (DWPI)をご利用いただくことをお勧めいたします。

DWPIは、世界中の特許を統一した言語(英語)で作成し、一貫性をもった基準で作成されています。

また、多国出願による同種のコンテンツの重複も避けることができます。

その解析結果の精度は、より優れたものとなり、解析の角度もより多彩となります。

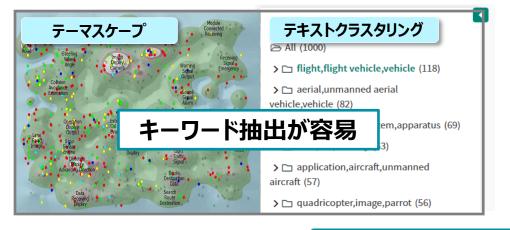
Derwent Innovation 上のテーマスケープ(ThemeScape)、テキストクラスタリングはDWPIのコンテンツをシームレスに解析できます。

## DWPI抄録を活用したテキストマイニング

※アナリスト契約のみ

例えば、おむつ関連技術の検索結果を 通常の抄録とDWPI抄録(用途)を 用いてテーマスケープを作成した場合

通常の抄録



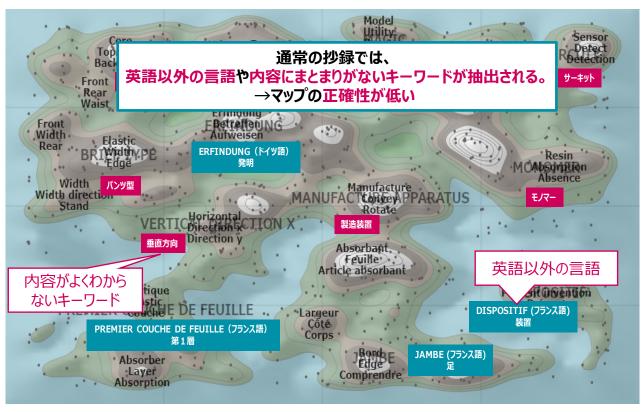




DWPIの特徴(統一言語、 観点ごとに整理された抄録、 ファミリ単位など)を最大限活 かした、精度の高いマクロ分 析が可能



#### 各項目のワードをテキストマイニングすることで有用なデータ解析が可能

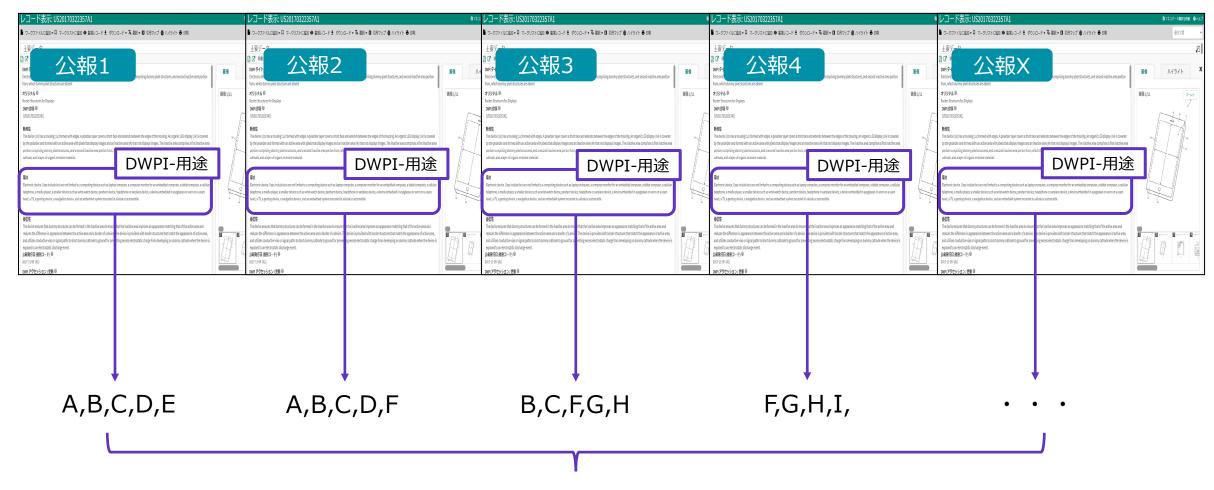




### Derwent Innovationの言語解析処理のイメージ①

例えば、作成した母集団で言語解析処理を行うテキストとして"DWPI-用途"を選択した場合、 母集団の各公報の"DWPI-用途"のテキストからワードを抽出します。

(解析する対象は公報レベル、ファミリーレベルで選択できます。DWPIフィールドはファミリー毎の作成になるので、DWPI抄録を解析する場合は、ファミリーレベルをおすすめします。)



**⇔** Clarivate

これらのワードが分析対象となります (テーマスケープでは、解析不要なワードの、ストップワード設定が可能です)

### Derwent Innovationの言語解析処理のイメージ②

各公報の"DWPI-用途"のテキストから抽出した ワードから公報同士で一致するワード数を見ます。

特許公報

шШОІС	1寸5十八十八					
を見ます。	公報1	公報2	公報3	公報4	•••	
ワード	A,B,C,D,E	A,B,C,D,F	B,C,F,G,H	F,G,H,I	• • •	
A,B,C,D,E	-	4	2	0		
A,B,C,D,F	4	ı	3	1		
B,C,F,G,H	2	3	-	3		
F,G,H,I	0	1	3	-		

性芸士/八芸品

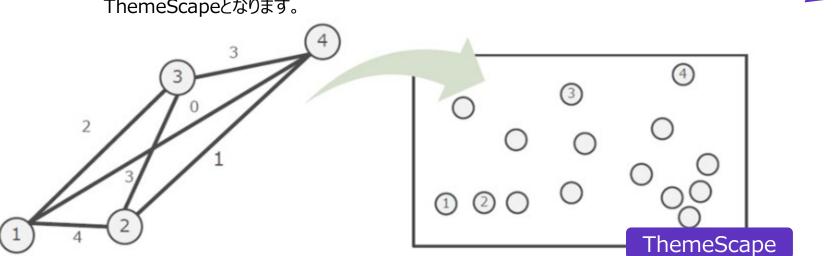
一致するワード数で距離を表現し、2次元化したものが ThemeScapeとなります。

公報1

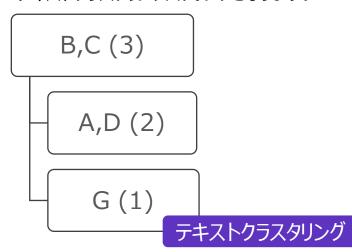
公報2

公報3

公報4



一致するワードをツリー状で表現したものが テキストクラスタリングスタリングとなります。



### **Derwent Innovationできるテキストマイニング**

DWPIコンテンツを活用 フォーカスを絞ったテキスト解析を 可能とする。

**テーマスケープ** 技術マッピングを行い全体俯瞰が可能

技術マップから個々のレコードや集合を抽出して、さらなるミクロな解析や活用が可能。

**テキストクラスタリング** 素早く、適切なキーワード抽出

スマートサーチの結果から のテキストクラスタリングもおすすめ。

→ 大量のテキストデータから、キーとなる単語やフレーズを発掘して、情報を類型化できる。 そこから得られる、マクロ情報を俯瞰して、整理、抽出して、個々の技術に触れていける。

Derwent Innovationでは、テキストマイニング機能としてテキストクラスタリングとテーマスケープを搭載、 **簡単**な手順・操作で**直感的にテキストマイニング**を実施することを可能とします。



# テーマスケープ描画における コツやワザ紹介を紹介



## ThemeScape描画におけるコツやワザ

ThemeScapeマップを使う上での悩み

思ったような技術マップが書けない。(基本的に位置は操作できないが対象を調整)

技術マップから読める傾向が伝わらない。(説明が必要?)

• マップ全体像の調整

- フォーカスを絞る

- ノイズを除去する

傾向を際立たせるマーキング

フィールド項目のマーキング (グループ色づけ機能)

- 自社情報を使ったマーキング

【切り出し】

【ストップワード】

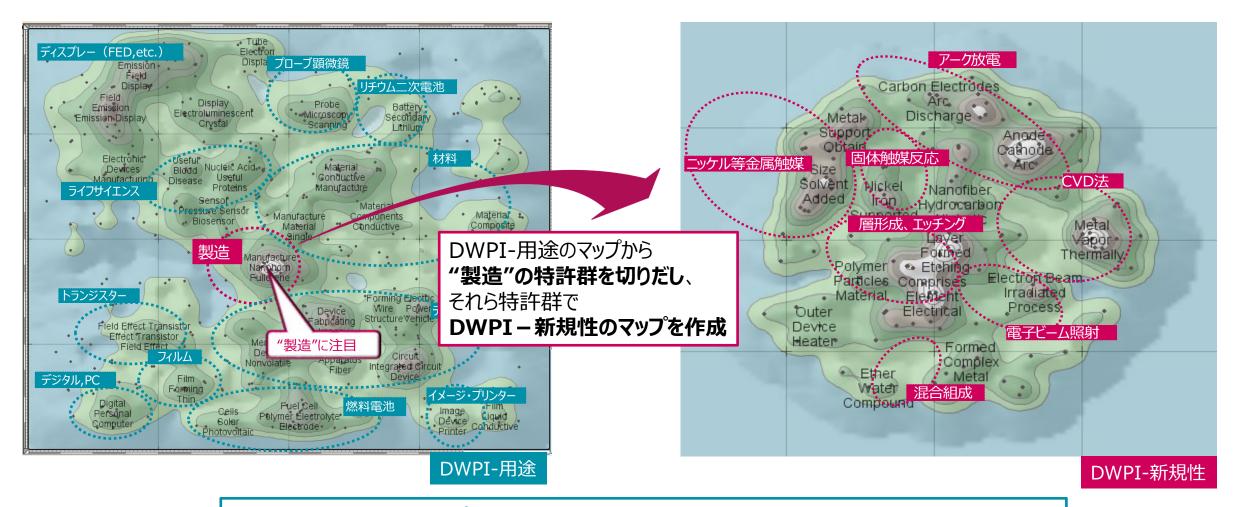
【色づけ】

【カスタムフィールド】

より意図が伝わる、魅せる ThemeScape マップを作りましょう。



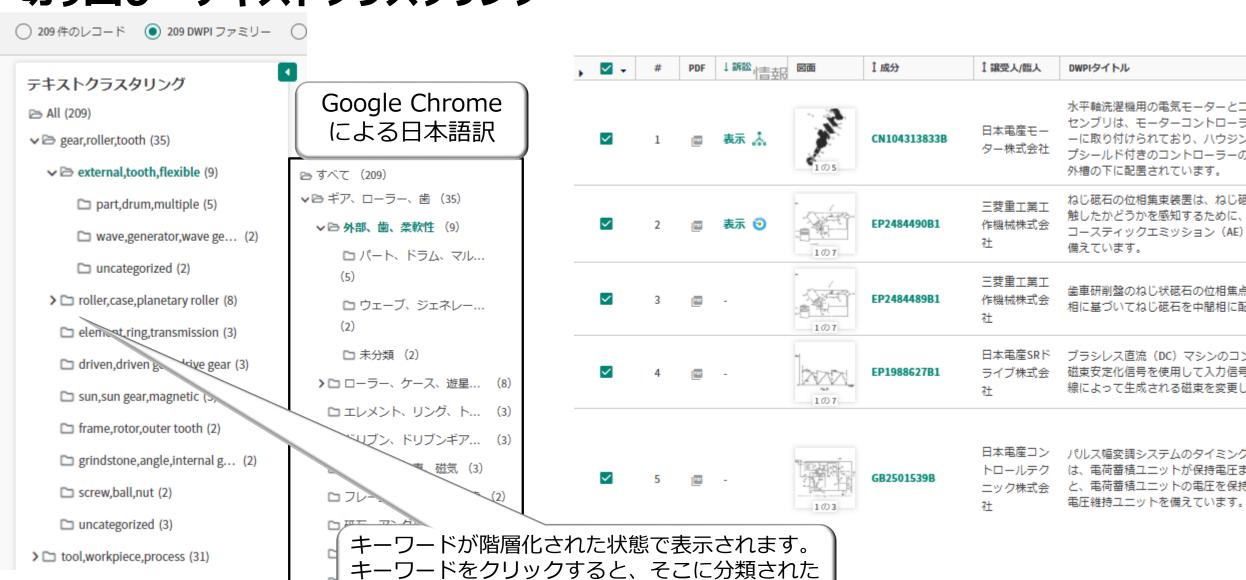
## 全体概要から一部の分野の詳細を探る





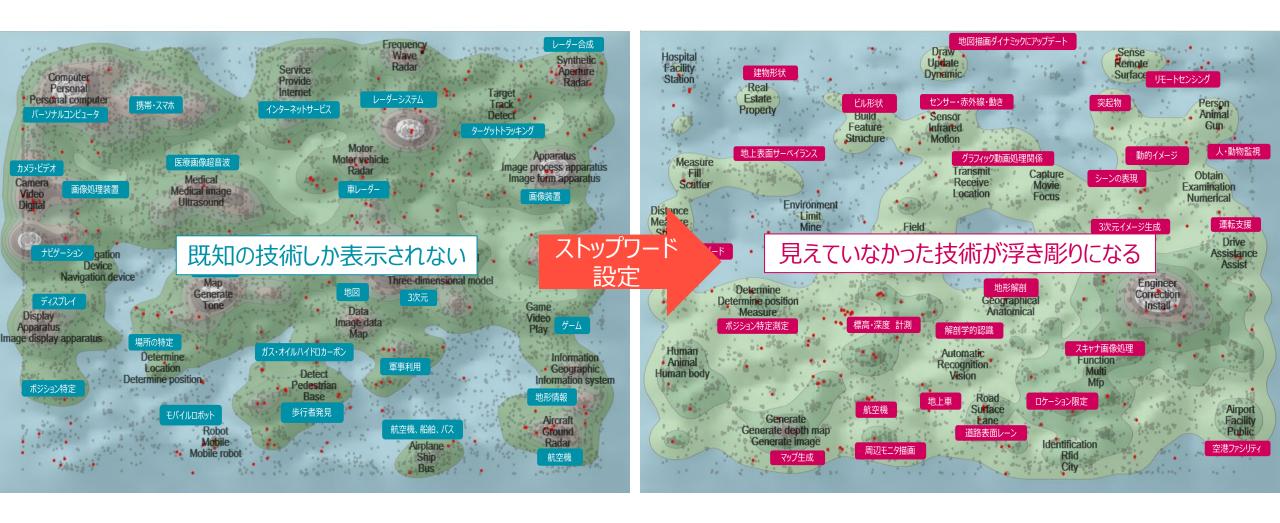
マップから興味のあるワードの特許群を切りだし、 それら特許群で新たなマップを作成することで、より詳細な技術内容の把握ができます。

### 切り出し→テキストクラスタリング



レコードのリストが表示されます。

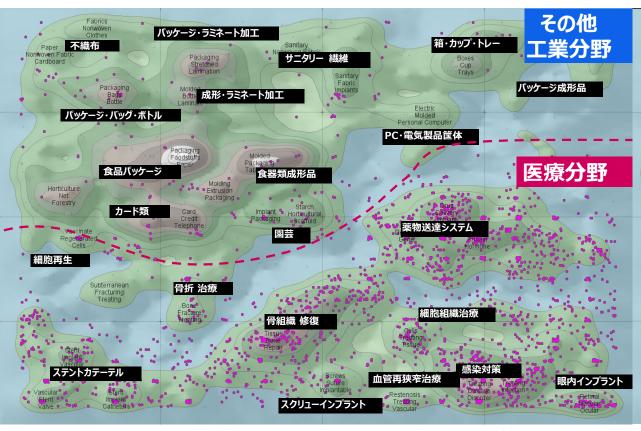
### 新規技術を見出す



マップから頻出のワードを**ストップワードとして設定**し、再度マップを作成することで、 埋もれていた**新たなワードを把握**することができます。



### 動向を知る(1)



箱・カップ・トレ-

OWPIクラス:B

※DWPIクラスB = 医薬: 医薬、獣医薬及びそれらの中間体

例:技術分野の把握

例:年代別の技術傾向の把握

2005-2010年2000-2004年1995-1999年1990-1994年



技術分類別、年代別、出願人別などのフィールド項目でマップに色付けをすると、 傾向の把握が容易にできます。

### 動向を知る(2)

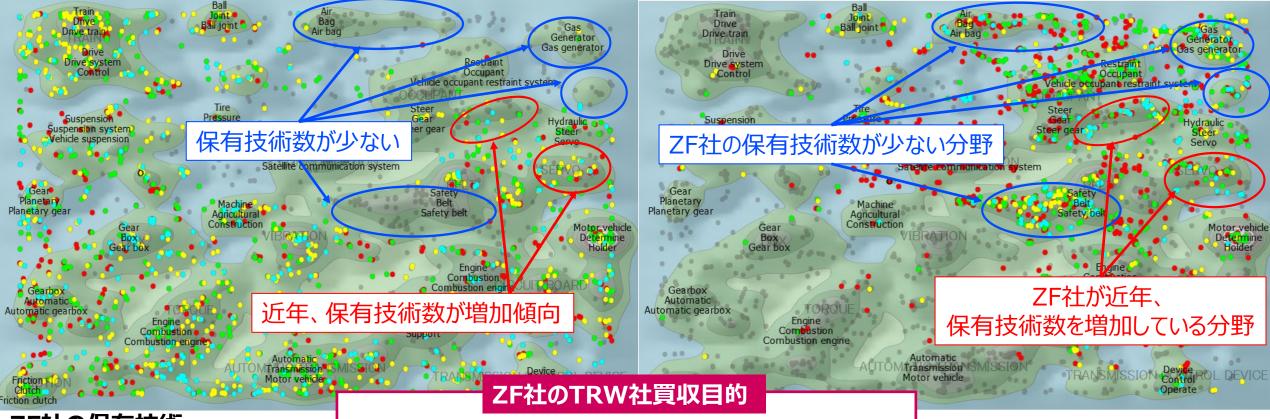
ZF社のTRW社買収の目的を考察するために、

ZF社、TRW社、2社の技術でテーマスケープを作成し、企業ごとで年代別の色付けをする

#### 日本經濟新聞

独自動車部品ZF、米同業TRWを1兆4400億円で買収 デンソー抜き業界3位に

2014/9/16 2:41



### ZF社の保有技術

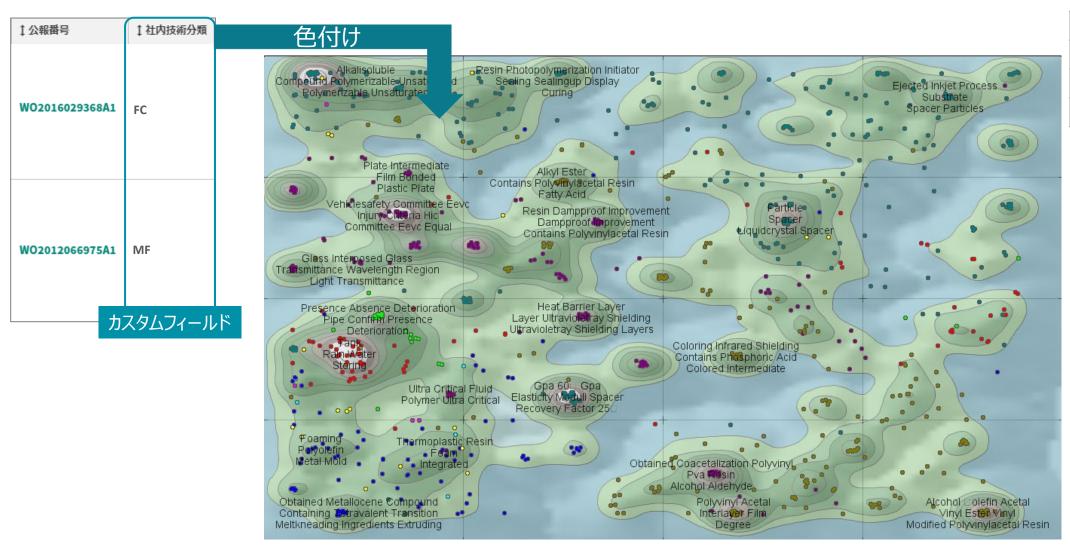
テーマスケープからZF社は、自社技術で手薄なところを TRW社の保有技術で補完する目的でTRW社を買収したと思料

### TRW社の保有技術

○ 2012-2015年○ 2008-2011年○ 2004-2007年○ 2000-2003年

マップに色付けすることで、自社・他社の保有技術の強み、弱みを把握し、 M&Aや提携先の検討もできます。

### 動向を知る(3)



【部門コード】

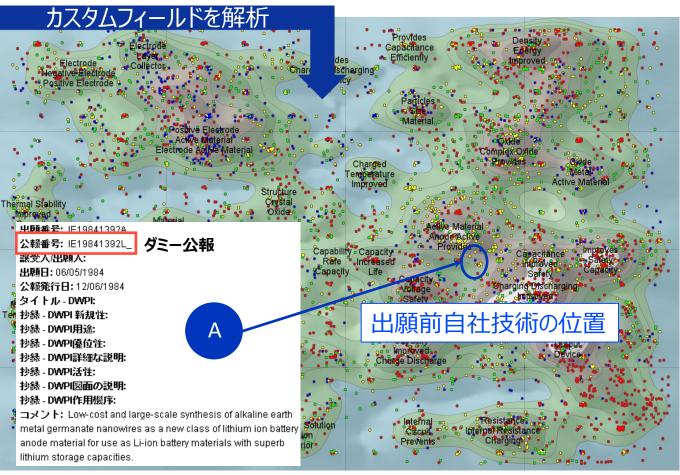
FC
 FP
 MF
 PL
 IT
 SL
 FM
 HH

社内技術分類などを入力するカスタムフィールドを設定し、

自社保有技術のテーマスケープでカスタムフィールドを色付けすると自社技術のポートフォリオを把握することができます。

### 出願前の自社技術の位置づけを探る





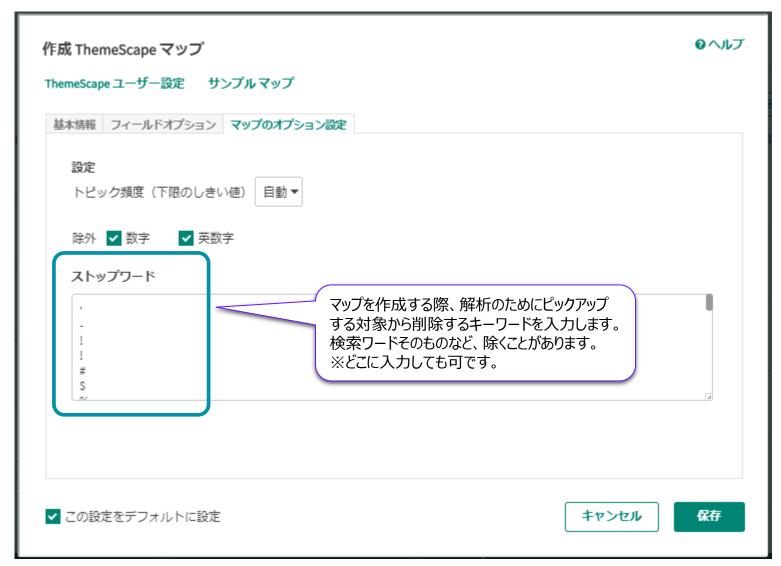
技術内容を入力するカスタムフィールドを設定し、

技術の母集団の中でダミーになる公報をピックアップし、ダミー公報のカスタムフィールドに自社技術内容を入力し、 テーマスケープでカスタムフィールドを解析すると自社技術の位置を把握することができます。

# -ストップワードの設定方法-



### ストップワードの設定方法:「マップのオプション設定」





### ストップワードの選び方

- ◆ DWPI特有のワードを除く
- ◆ 母集合作成に使用した検索キーワードを除く
- ◆ 母集合の特徴・分野に特有のワードを除く
- ◆ 既存マップにストップワードを追加して作成し直す手順



#### DWPI特有のワードを除く

DWPI抄録でよく使われているが、分析に必要な技術的な意味がないワードがあります。

- ◆ DWPI抄録 優位性に頻出ワード: excellent
- ◆ DWPI抄録 用途に頻出ワード:application



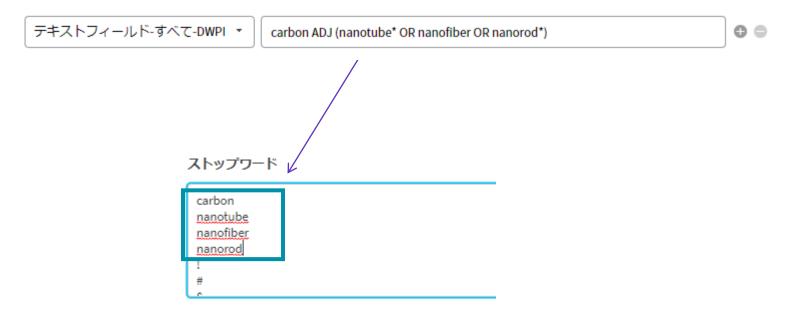


#### 母集合作成に使用した検索キーワードを除く

検索キーワードはマップ上に沢山表示されることがあります。これにより新しい気付き (想定外のワード)が埋もれてしまう可能性があります。既に把握している情報のため、 分析対象から外しても良いかもしれません。

#### 検索フィールド

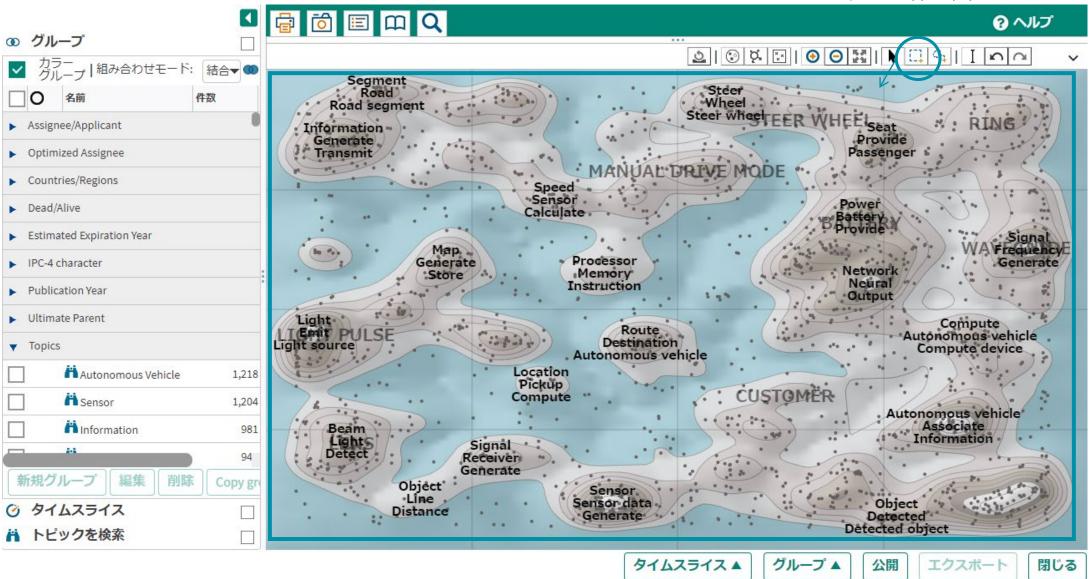
フィールドや演算子 (AND、OR、NOT) を組み合わせて検索条件を作成します。ヘルプが必要な場合、クエリー作成の基本 Z を表示するか、または選択メニューの中のフィールドの詳細を参照してください。



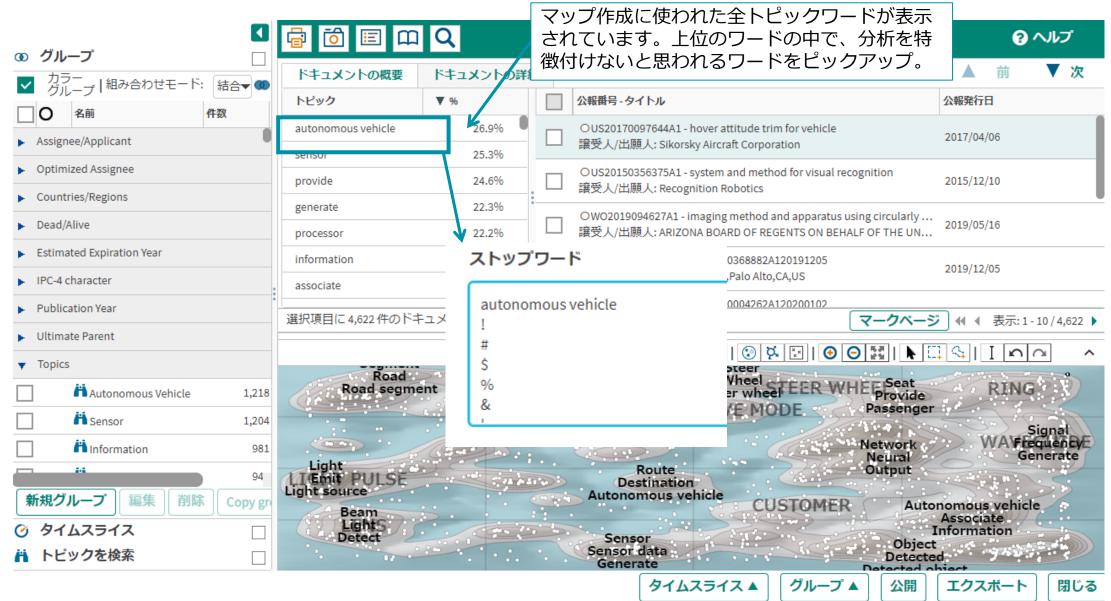


### 母集合の特徴・分野に特有のワードを除く①

#### マップ全体を囲む

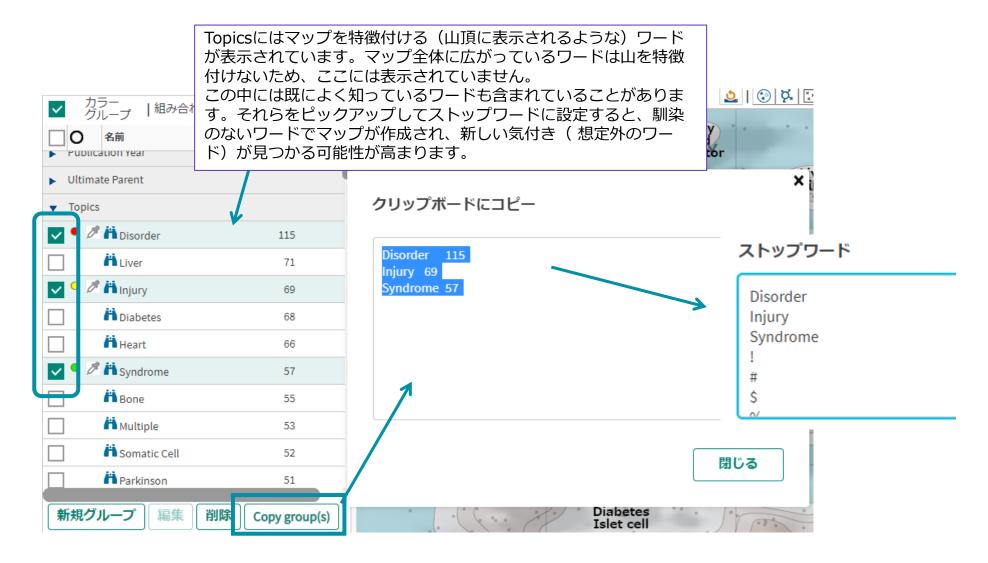


### 母集合の特徴・分野に特有のワードを除く②



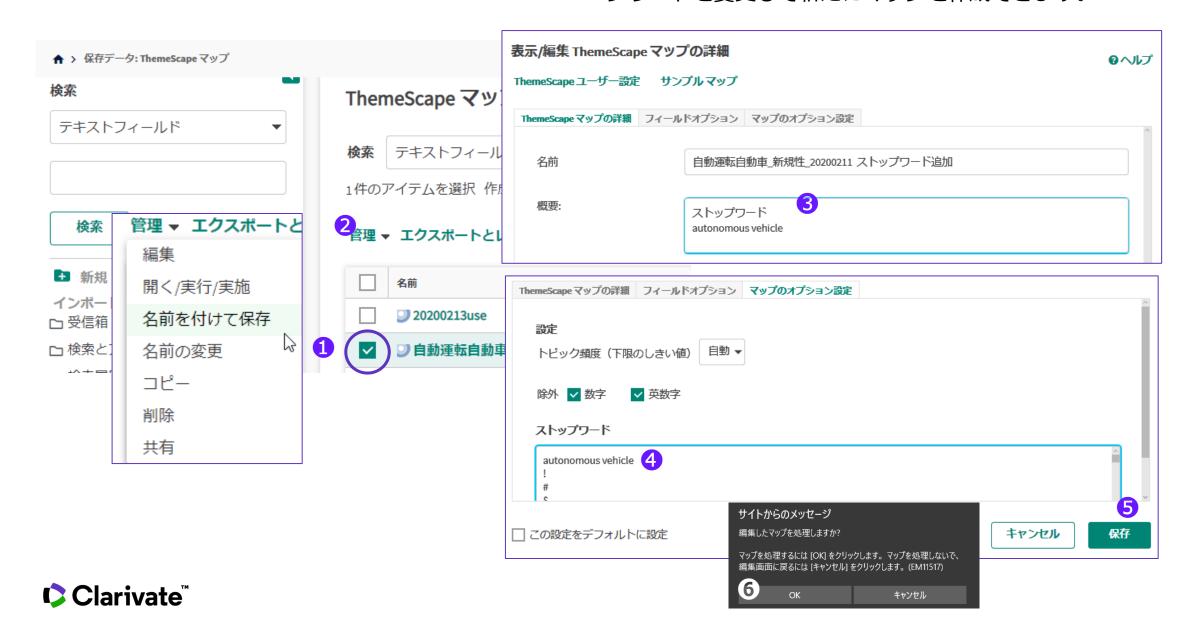


#### 母集合の特徴・分野に特有のワードを除く③





# **既存マップにストップワードを追加して作成し直す**の手順ですと、元のマップを残したまま、ストップワードを変更して新たにマップを作成できます。



# -カスタムフィールドの活用・ダミー公報-



### 設定方法①-管理者によるカスタムフィールドの作成





設定方法②-テキストフィールドを有するカスタムフィールドの作成

"カスタムフィールドの作成"画面が開きます。

"基本情報"のシートの"名前"に使用するフィールド名を入力します。

※フィールド名は、32文字(かな文字も可能)までです。

"形式"でどのようなタイプのフィールドにするか下記の中から一つ選択します。

- ・日付フィールド
- 複数選択フィールド(複数選択リストまたはチェックボックス)
- ・数値フィールド(%記号とともに表示)
- 単一選択フィールド(ドロップダウンリストまたはチェックボックス)
- ・テキストフィールド ← 2048文字まで入力可能です。
- ・URLフィールド
- 年フィールド
- ※複数選択フィールド、単一選択フィールドを選択すると"表示方法"の選択が必要となります。

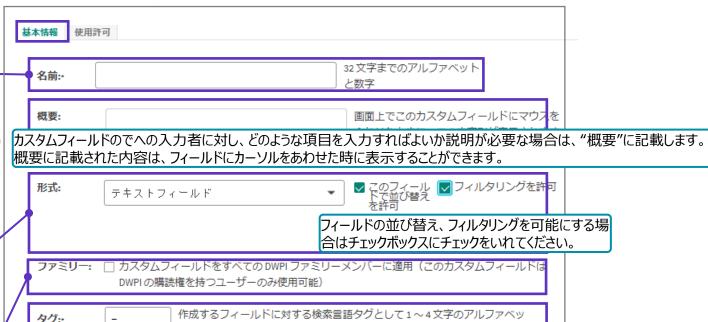
DWPIファミリーメンバーに対し、全て同じ値を設定するオプションです。 作成したフィールドをファミリー単位で使用する場合は、"ファミリー"のところに チェックをいれます。

英数字で1~4文字でタグを設定します。

タグ設定によりエキスパート検索でカスタムフィールドも識別子になり、検索フィールドとして使用できます。

⑤でファミリーにチェックをした場合、タグの先頭は+になります。設定しなかった場合は、タグの先頭は-となります。

6



#### フィールドタイプ

す (例: ab12)
フィールドのガイドライン: このフィールドタイプは2048 文字までです

タイプ	文字数	説明	活用例
ドロップダウンリスト (カラーフラグ可)	64文字 (全角半角)	複数リストの中から1つ選択する形式 (リスト上限:1000)	社内分類、部署、製品名
ラジオボタン (カラーフラグ可)	64文字 (全角半角)	複数リストの中から1つ選択する形式	特許区分:自社・受入・譲渡済 審査ステータス
複数選択ボックス (カラーフラグ可)	64文字 (全角半角)	複数リストの中から複数選択する形式 (リスト上限: 1000)	社内分類
チェックボックス (カラーフラグ可)	64文字 (全角半角)	複数リストから複数選択する形式	社内分類
テキスト	2048文字 (全角半角)	フリーメモ	特許に関するコメント 製品名
日付	YYYY-MM-DD	日付形式 - 年月日	年金更新期限 特許存続期限
数値	15桁	数值	年金更新金額
年	YYYY	年	
URL	2048文字 (全角半角)	Web, 社内サーバー等へのURL	製品のURL 関連資料へのURL

トや数字を入力・四角の中のアンダーパーは、タグの必須プリフィックスで



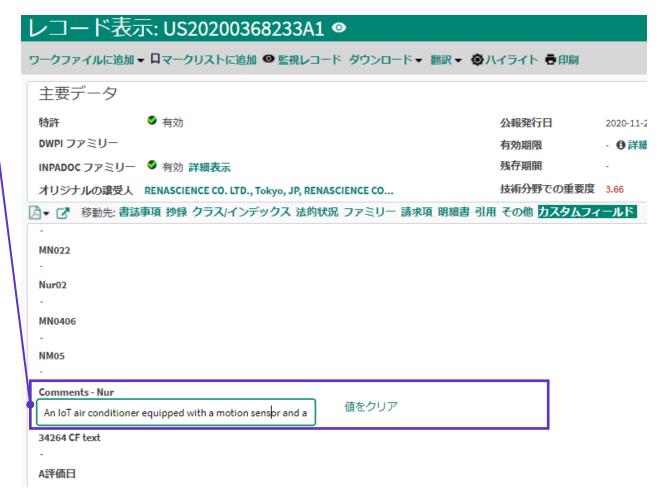
## ダミー公報を含むThemeScapeの描画

①カスタムフィールドでのダミーレコードの設定

ダミー公報を決めて、ダミーに用いるカスタムフールド (「テキスト形式」に限る、例では「Commentー Nur」)に、技術的内容を入力する

入力例: An IoT air conditioner equipped with a motion sensor and a brightness meter that determines whether a person is sleeping and adjusts the temperature.

※「管理者」または、用いるカスタムフィールドの「読み取りと編集権限のあるユーザー」である必要があります。 ※用いるカスタムフィールドは形式が「テキストフィールト、」である必要があります。





# ダミー公報を含むThemeScapeの描画

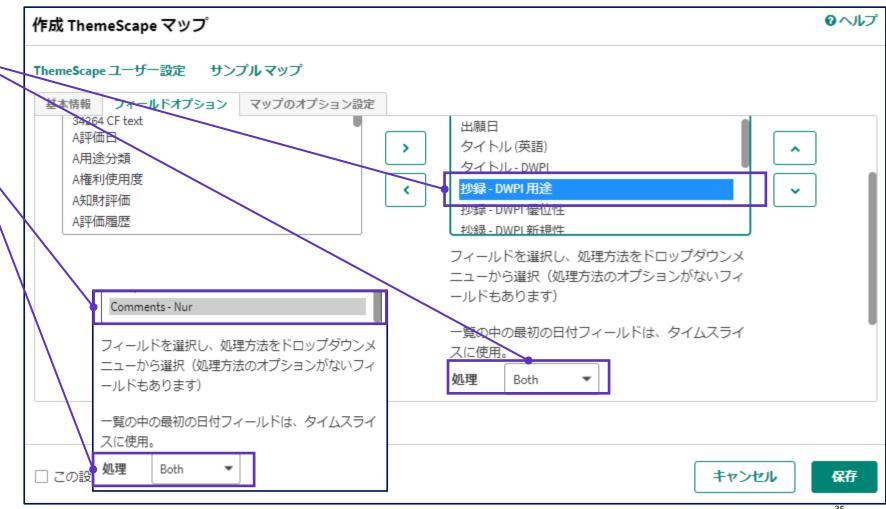
②解析対象に用いたカスタムフィールドを追加してThemeScapeを描画する。

主として解析対象とするフィールドを設定。右では用途マップ。

用いたカスタムフィールド(例で「Comment-Nur」)も解析対象に設定

※「管理者」または、当該カスタムフィールドの「読み取りと編集権限のあるユーザー」である必要があります。





# おすめ動画紹介 今後のwebセミナー



### おすすめwebセミナー動画紹介①

テーマスケープとテキストクラスタリングの具体的な活用法を紹介する動画です。

新たな事業チャンスを見出す (2020/9/2開催)

テーマスケープテキストクラスタリングを活用 した新規用途探索を具体的な事例紹介

URL:https://interest.clarivate.jp/202103 expolore-use webinar dw



**Clarivate** Derwentオンデマンドセミナー: https://clarivate.com/ja/derwent-ondemand-webinars/

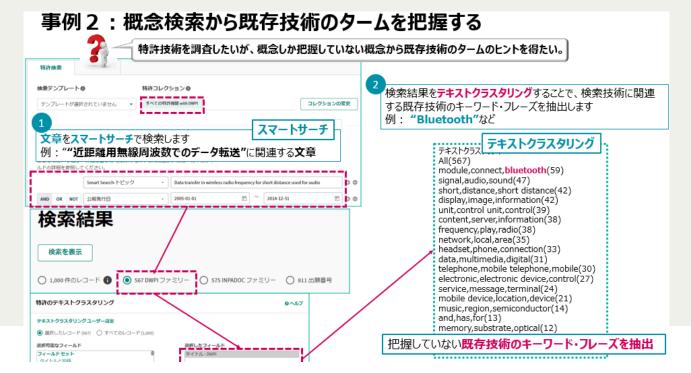
### おすすめwebセミナー動画紹介②

テーマスケープとテキストクラスタリングの具体的な活用法を紹介する動画です。

• 関連キーワードを効率的に引き出す (2020/10/15開催)

スマートサーチ、テキストクラスタリングを 活用したキーワードの抽出とデータの俯瞰

URL:https://interest.clarivate.jp/202103 related-keywords webinar dw





### Derwentオンデマンドセミナーへのアクセス

TOP Page (日本語サイト): https://clarivate.com/ja 🛑 🚉 🚛



Clarivate TOP PAGEからのアクセス

Clarivate\*

会社概要

ブランド・

イベント お問い合わせ

①クラリベイトのHPの画面上"イベント"をクリックします。



#### 人間の創意工夫が世界を変える

インスピレーションはふとした瞬間に生まれ、イノベーションのライ フサイクルが始まります。

クラリベイトがほんの少し背中を押すことで、漠然としていたアイデ アが暮らしを変えるイノベーションになります。



#### Derwent オンデマンドウェブセミナー

様々なテーマで知財情報の活用についてご紹介しています。日々の業務 にお役立てください。

②イベントセミナー情報の "Derwentオンデマンドウェブセミナー" を探して「ボタン」をクリックします。



### 過去のイベント

#### Derwent オンデマンド ウェブセミナー



関連キーワードを効率的に引き出す -スマートサーチ、テキストクラスタリングを 活用したキーワードの抽出とデータの俯瞰

本セミナーでは、スマートサーチを用いて抽出した検索結果をテキスト クラスタリングでワード分類することで、キーワードを簡単に抽出する 方法を紹介し、技術への把握を効率的に行うことを提案します。



#### 新たな事業チャンスを見出す 用途探索

められています。 本セミナーでは、Themescape等を使い、自社保有技術 を活かした用途探索の仕方をご紹介します。

詳細はこちら





### お問い合わせ

#### サービス全般に関するお問い合わせ

Tel:03-4589-3101 Fax:03-4589-3240

〒107-6118 東京都港区赤坂5丁目2番20号 赤坂パークビル18階

<u>カスタマーサービス(ヘルプデスク</u>) Tel (フリーコール): 0800-170-5577 Tel: 03-4589-3107

Email: <u>ts.support.jp@clarivate.com</u>

サービス時間:

土・日・祭日を除く午前9時30分~午後5時30分

