

俯瞰解析ソフトの比較

PLASDOC協議会

- JXリサーチ株式会社 新井(発表者)
- JNC株式会社 谷岡
- 出光興産株式会社 田村
- 昭和電工株式会社 皆川
- 株式会社カネカテクノロジー 矢野

【はじめに】

本発表は、2017年度PLASDOC協議会WG活動で検討した内容を、一部変更して発表させていただきます。

【日本PLASDOC協議会(JPA)】

- 設立 1970年4月
- 会員数 23社(2018/4/1現在)
- 高分子(ポリマー)を中心にしたオープンな情報検索研究会で、これまでの経緯から化学系企業が中心ですが、現在では**業種の枠に捉われない研究会**となっております。
- 活動内容
毎年4~5つの**ワーキングテーマ**を設定してワーキンググループを作り、**約半年間の活動**を行います。顔を合わせたの議論以外に、メールでの意見交換も活発に行われます。また、グループによっては、昼食会やアフター5でも情報交換を行なって知見を広めています。
活動成果の発表は、**秋季研修会**(11月頃)で行います。研修会は宿泊をしての会議となるのが通常です。
- 参加申し込み、お問い合わせは下記ホームページを参照ください！

<http://plasdoc.sakura.ne.jp/index.html>

【ワーキング活動の概要】

- 最近、キーワード抽出から俯瞰図を書かせるソフトが多く出回っている (ThemeScape、DocRadar、BizCruncher等)。それらの仕様・機能等の情報から特徴を比較した。
- さらに、それら俯瞰解析ソフトが生成する図にはどの様な違いがあり、山を形成している公報の集合にどのような差があるのか比較した。



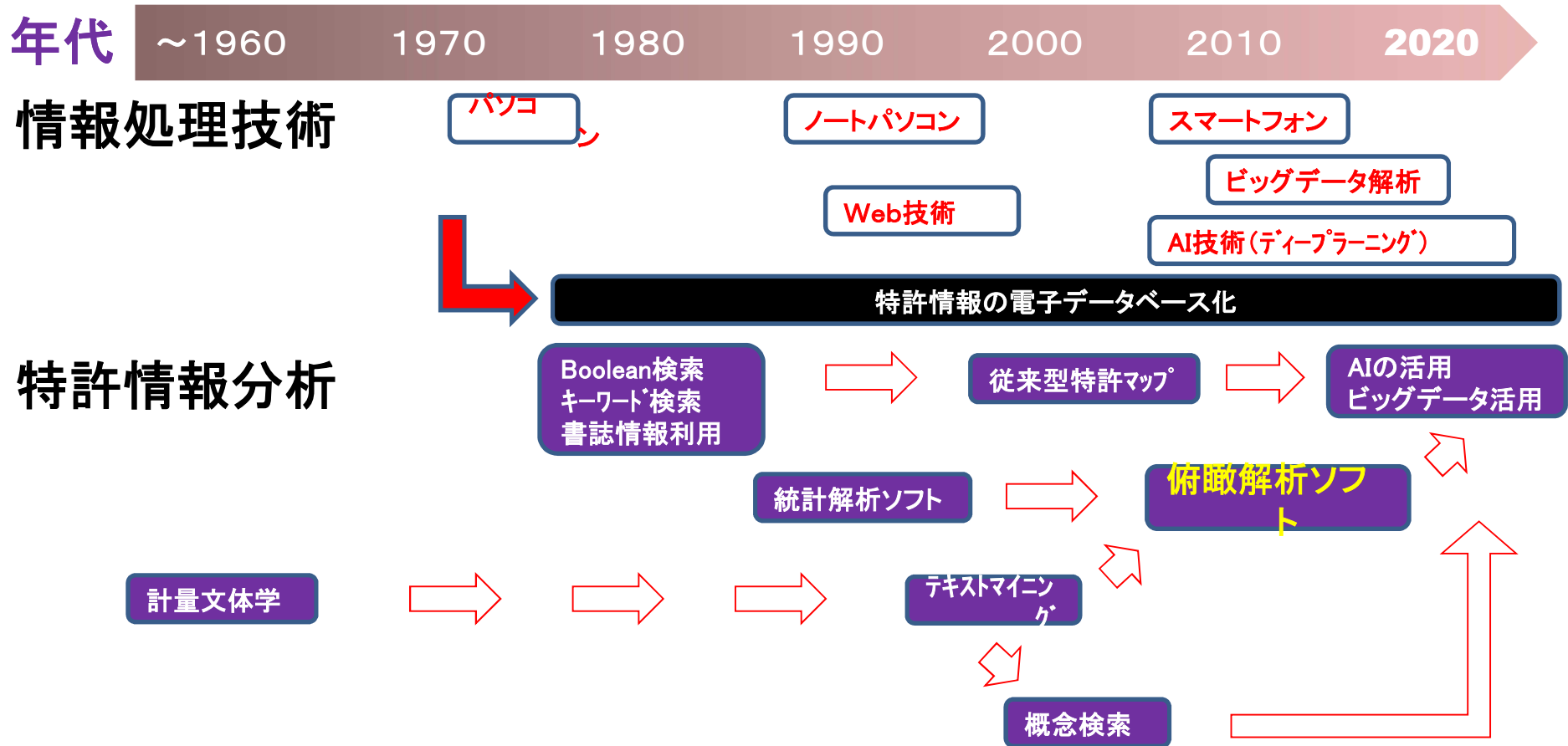
【検討項目】

- 俯瞰解析ソフトの現状
- ThemeScapeの原理、特徴
- DocRadarの原理、特徴
- DocRadarとThemeScapeの比較

【検討詳細】

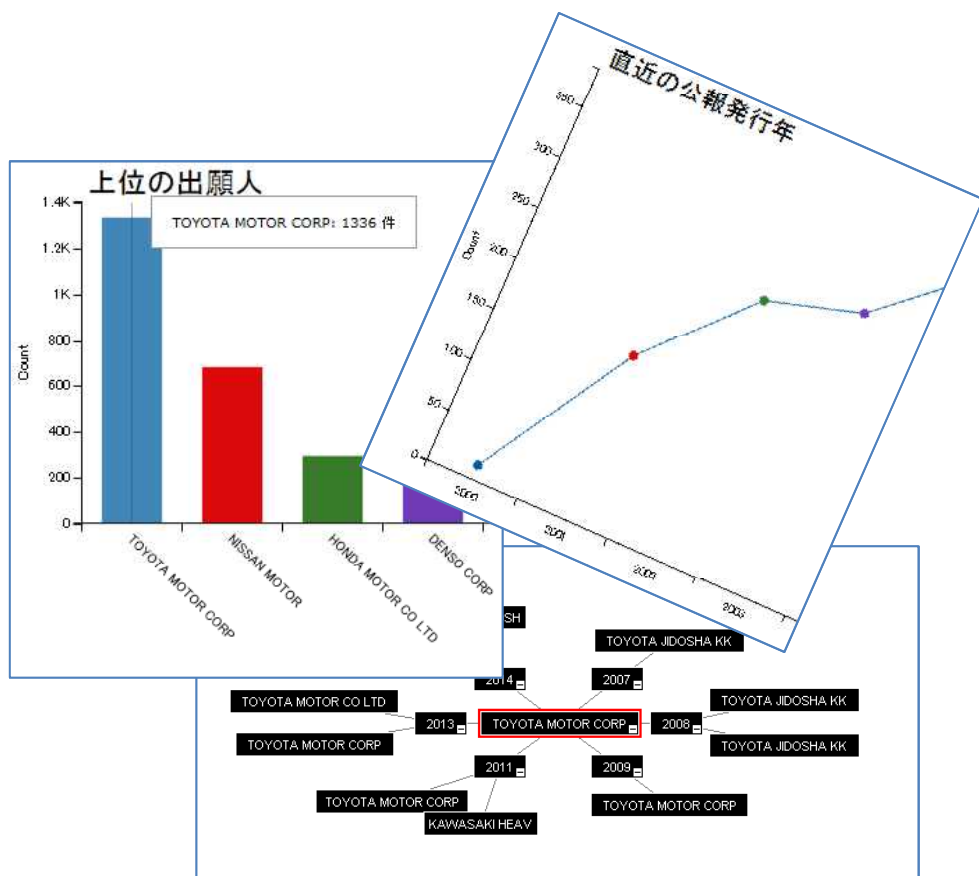
- 俯瞰解析ソフトの現状
- ThemeScapeの原理、特徴
- DocRadarの原理、特徴
- DocRadarとThemeScapeの比較

■ 情報処理技術と特許情報分析の進歩

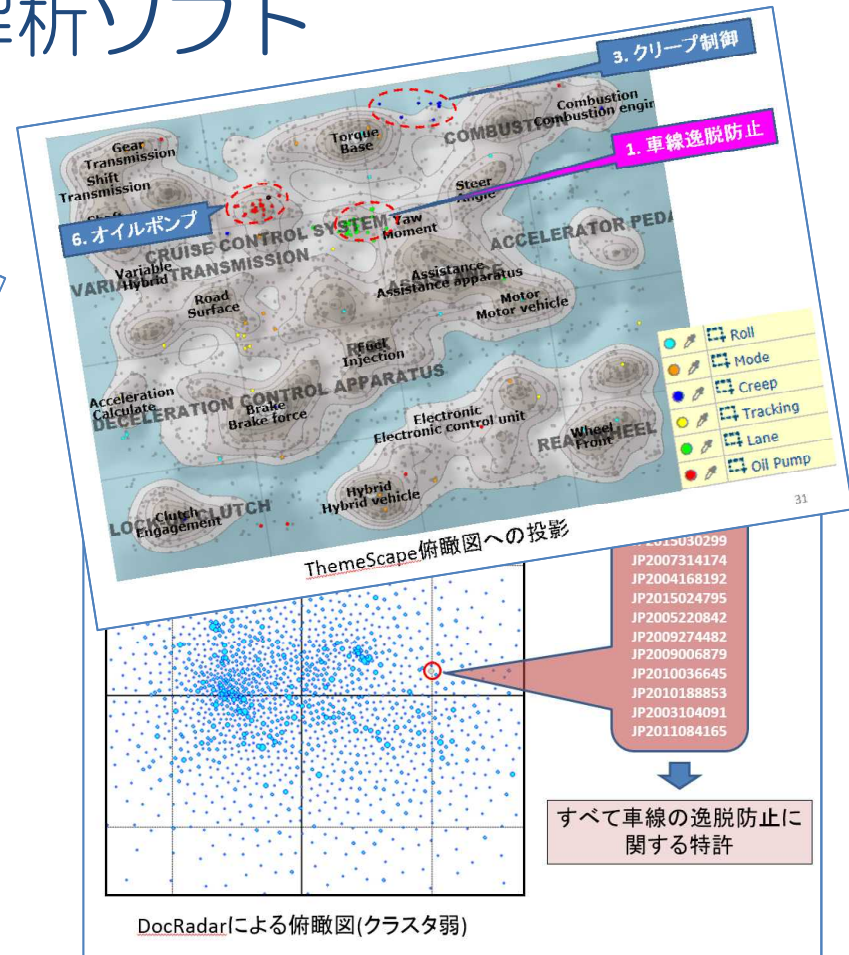


- 特許情報分析の技術進歩は激しく、古くからあるBoolean検索が主流ではあるが、概念検索に應用されているテキストマイニング技術の進歩が著しい。
- 最近では、統計解析とテキストマイニング技術を應用した、俯瞰解析ソフトの利用が活発になっている。

■ 従来型特許マップと俯瞰解析ソフト



従来型特許マップの例



俯瞰解析ソフトの例

- 従来型特許マップは、主に明細書の書誌事項を基に、出願人分析や発明者分析、特許分類分析、引用情報分析などを行う。
- 俯瞰解析ソフトは、明細書中の文章にテキストマイニング技術を適用・俯瞰可視化して、分野や類似性、傾向、関連性などを分析する。 8

■ 俯瞰解析ソフトおよび従来型特許マップの例

	俯瞰解析ソフトの例	メーカー	URL
1	ThemeScape (Derwent Innovaton)	Clarivate Analytics	http://clarivate.jp/products/ti/
2	DocRadar、TechRadar	VALUENEX	https://valuenex.com/
3	Biz Cruncher	パテントリザルト	http://www.bizcruncher.com/
4	TRUE TELLER パテントポートフォリオ	NRIサイバーパテント	http://www.patent.ne.jp/service/macro/gaiyo.html
5	Patent Integration	Patent Integration	https://patent-i.com/ja/index/
6	Orbit Intelligence	Questel	http://www.questel.jp/orbit-intelligence.php
	従来型特許マップの例	メーカー	URL
7	パテントマップEXZ	インパテック	https://www.inpatec.co.jp/software/patentmap
8	ぱっとマイニングJP	JPDS	http://www.jpds.co.jp/map/index.html
9	PAT-LIST-GLS	レイテック	http://www.raytec.co.jp/products/index.htm
10	テクノマップ	アルトリサーチ	http://www.patentsearch-japan.com/
11	PatentGrid	DJ-SOFT	http://www.djsoft.co.jp/products/patentgrid4.html
12	PatentStrategies	LexisNexis	http://jp.lexisnexisip.com/
13	CsvAid	中央光学出版	https://www.cks.co.jp/home/product-2-2.htm
14	PatBase アナリティクス、VizPat 統計分析	RWSグループ	http://jp.rws.com/
15	HYPAT/DA	発明通信社	http://www.hatsumei.co.jp/

- 解析原理が近い、ThemeScape、DocRadar (TechRadar) について比較することとした。

【検討詳細】

- 俯瞰解析ソフトの現状
- ThemeScapeの原理、特徴
- DocRadarの原理、特徴
- DocRadarとThemeScapeの比較

■ ThemeScapeとは（開発元 Clarivate Analytics社）

- 大量の文書情報をテキストマイニングし、分野を俯瞰的に見るためのツール
*テキストマイニングは、大量の特許文献の書データを様々な観点から分析し、役に立つ知識や情報を取り出し、利する技術技術。
- 文書データが等高線マップで表現され、距離・密度・分布の定量分析が可能。
- 分野の動向や年代傾向や出願の分野の分析、学術文献情報から最先端の傾向の把握、社・グループ企業のポートフォリオ分析、萌芽領域の解析。

○の距離が近い
＝類似性が高い

共通のトピック（特徴語）を含むレコードが集まった「密集領域」は島や山で示される。

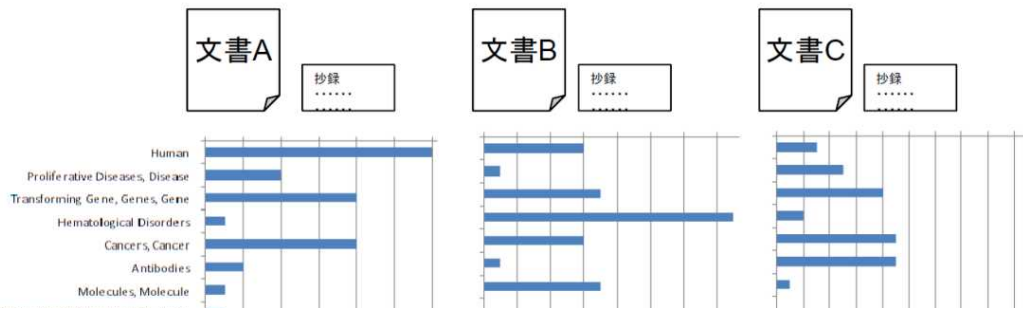


テキストマイニングなので、分析の元になるテキストデータの信頼性が一番大切！

専任スタッフが作成したDWPI抄録は英語の統制された単語が使用されており解析に有効！！

トピックの生成からマップができるまで

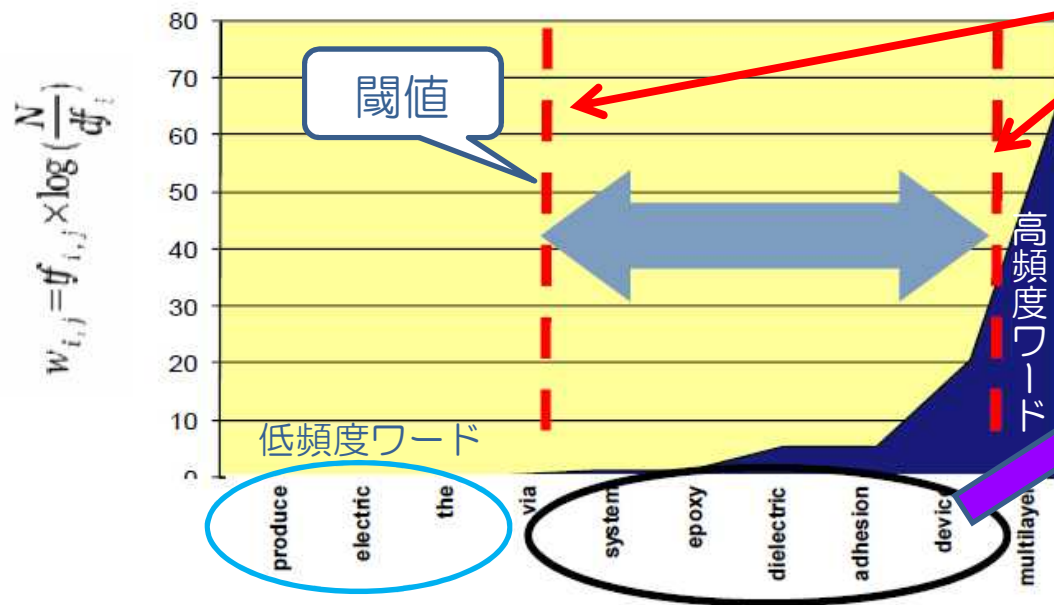
1. 文章を単語に分割し、単語リストを作成



2. TFIDF計算： Term Frequency/Inverse Document Frequency

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

3. TFIDFにより頻度/低頻度ワードを除外



どこまでの範囲の単語をトピックとして取り出すかの閾値は設定のところで選択することができる。

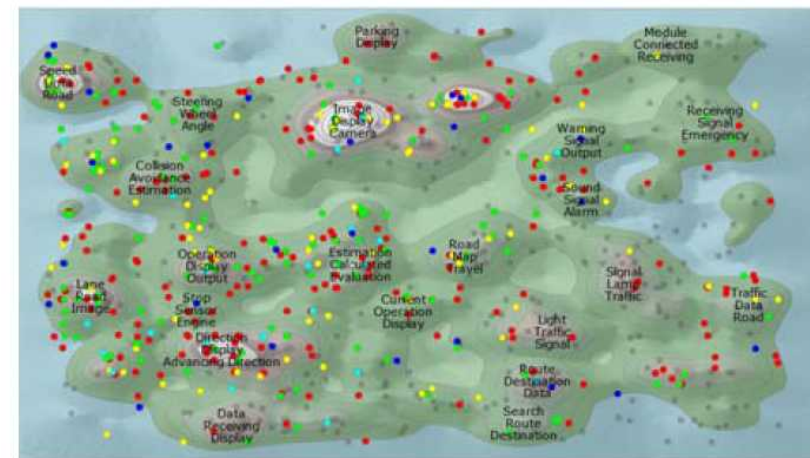
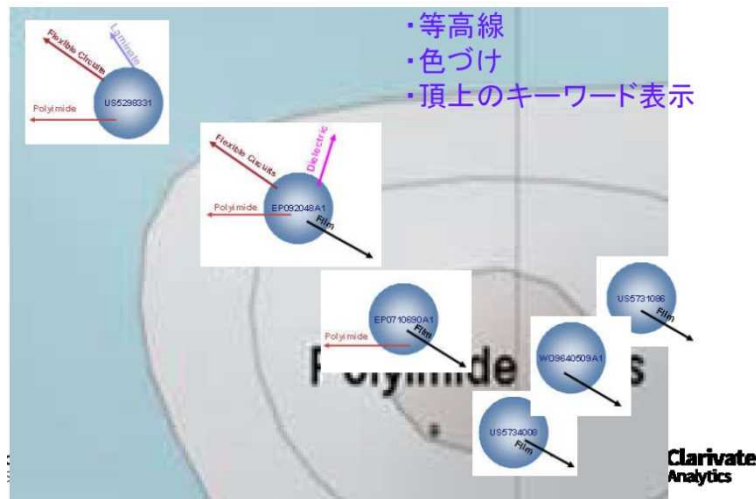
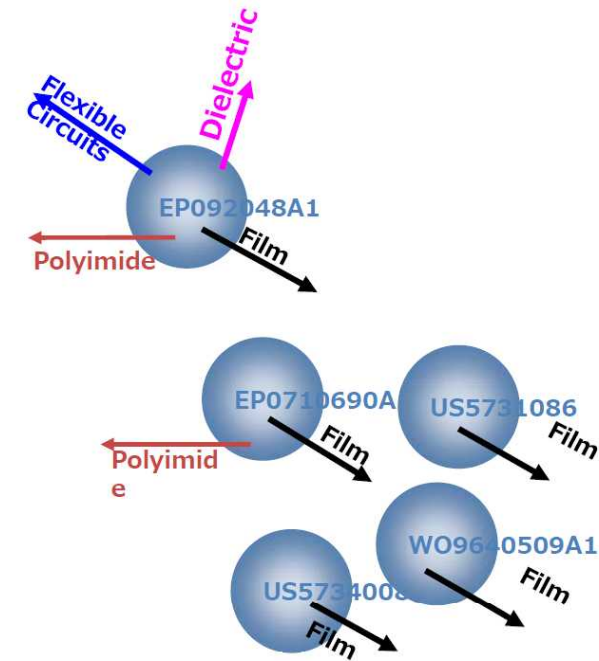
Topics	
Outputting, Output	.
Computing Device, Computer	.
Shaft Rotate, Rotating, Rotat...	.
Coupled	.

トピックとして使用されるキーワード

4.各文書のDocument Vectorsが付与される。

5. Vector Space Modeling (VSM) を適用 2次元上に特許文献をプロット。

6.等高線が記載されたマップに表示され、頂点とそれ以外の領域にトピックが表示される。



■ ThemeScapeにおける単語調整

1) 閾値調整…分析する単語(トピック)の閾値の設定

- マップに表示されるトピックの**閾値を選択可能**。
- 閾値を下げる → 低頻出の単語まで分析に使用可能。

表示/編集 ThemeScape マップの詳細

ThemeScape ユーザー設定 | サンプル マップ

ThemeScape マップの詳細 フィールドオプション マップ0

設定:

トピック頻度(下限のしきい値): 自動

除外: 数字 英数字

ストップワード:

excellent

!

#

\$

%

&

2) ストップワード(分析に使用しない単語)の設定

数字を含む単語はデフォルトでストップワード。

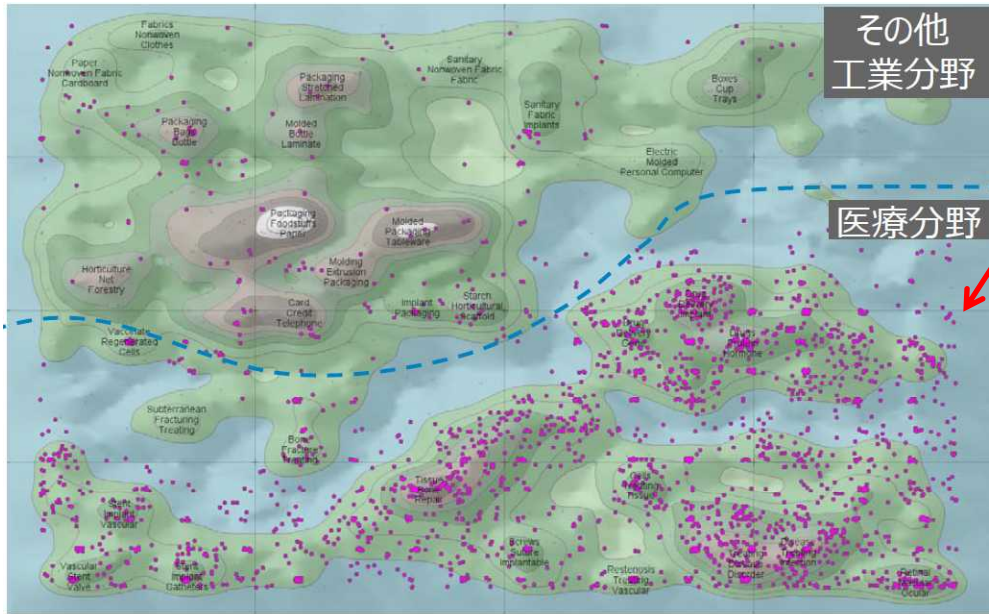
数字を含む単語を分析に使用する場合は設定解除。

デフォルトでストップワードが設定されているが個別に追加削除が可能。

- 特記事項：シソーラス設定できない。
複合語は別カウント。(例：oil と parm oil)

ThemeScape分析例

➤ 生分解性ポリマーに関する特許の分析例



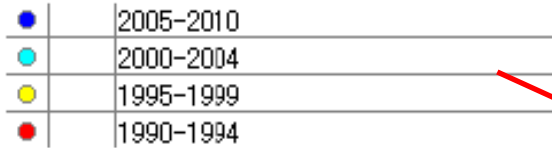
分野別に色付けした例



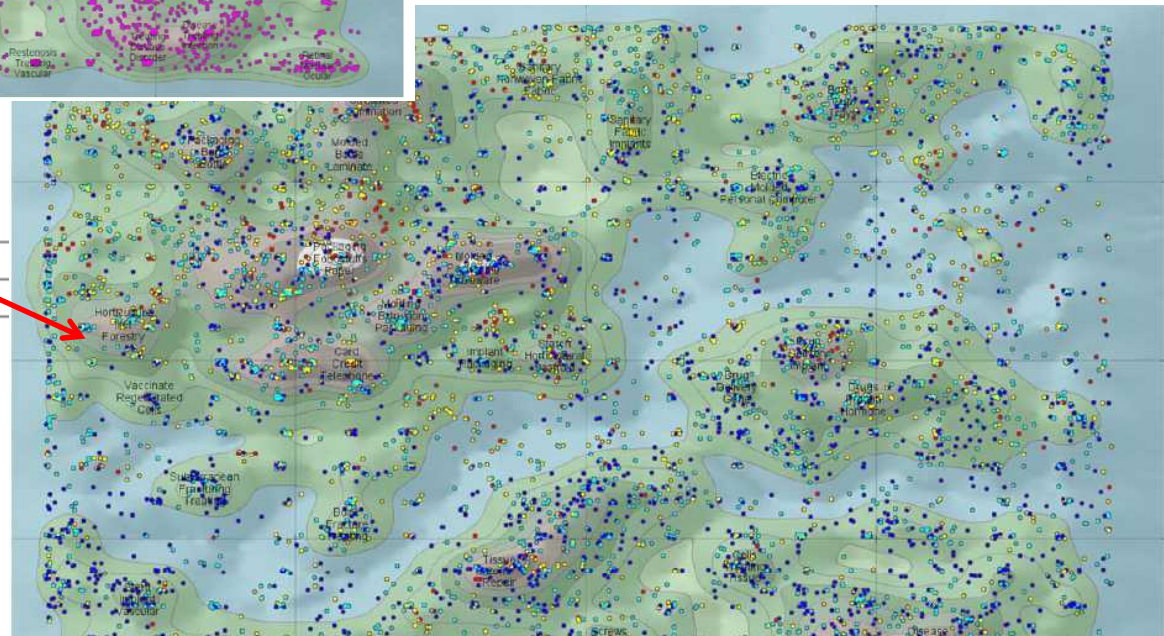
どのような技術が
どんな分野で多いか
を把握する。



➤ 時系列の動向の例



経時変化をみる。

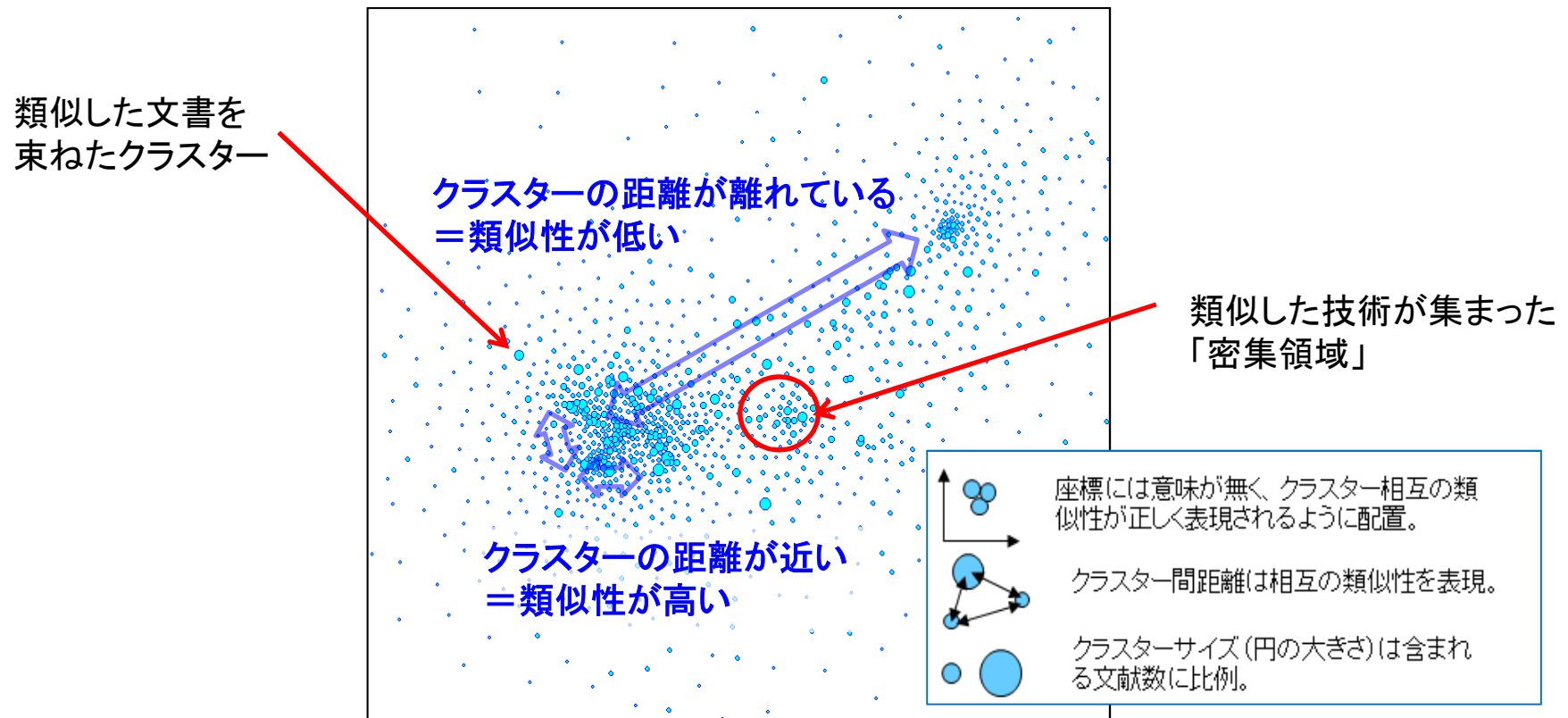


【検討詳細】

- 俯瞰解析ソフトの現状
- ThemeScapeの原理、特徴
- **DocRadarの原理、特徴**
- DocRadarとThemeScapeの比較

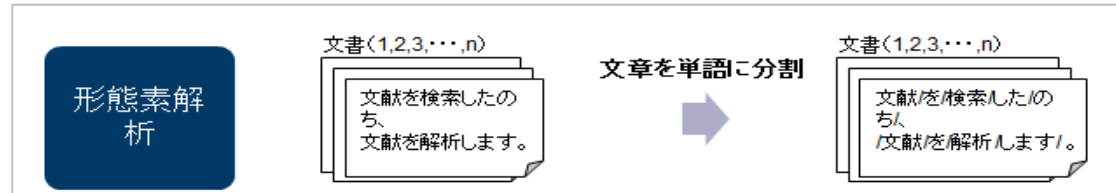
■ DocRadarとは (開発元 VALUENEX社)

- 大量の文書データを、類似度評価によって整理・クラスタリングし、俯瞰可視化する解析ソフト。
- 文書データに、X,Y座標情報が付与され、距離・密度・分布の定量分析が可能になる。
- 技術トレンドの把握、自社他社の注力技術や方向性の把握、萌芽領域、技術展開先の探索。
- 特許以外に論文やアンケートなど多様な文書データの解析・俯瞰も可能。



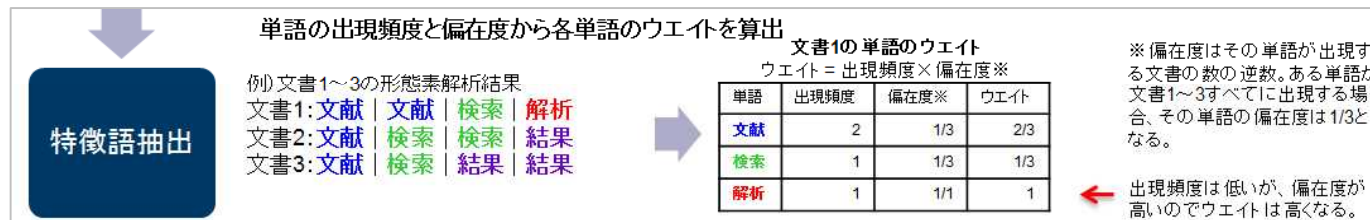
■ 特徴語の生成から座標が決まるまで

1. 分析申請時に重みづけをしたフィールドの文章を単語に分割し、単語リストを作成する。



2. 抽出された単語全件にシソーラスが適用される。

3. 各単語にウエイトが付与され、特徴語 (ThemeScapeというトピック) として抽出される。
(出現頻度×偏在度 + 単語調整の重みづけ)



4. 各文書の座標情報が付与され、クラスター解析が行われる

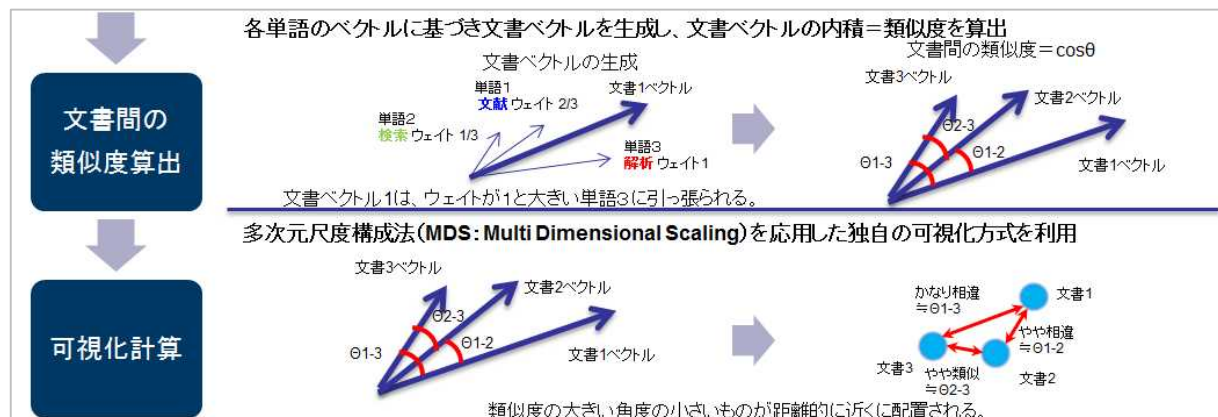


図: VALUENEX社の資料から転載

■ DocRadarにおける単語調整、シソーラス

● 単語調整

- 分析に使用した単語を一覧表で取り出すことが可能。
- 分析時の**ウエイト(重み)**を任意の値で設定可能。
- **ウエイトを” 0” にすると、禁止語として分析対象から除外。**
=ThemeScapeでいうストップワードと同じ処理になる。

● シソーラス (同義語辞書)

- 分析の際に**類義・同義語**とみなす単語の集合を定義可能。

● 特記事項

- 単語調整したリストを登録。
⇒同一分野なら、最初から単語調整済みリストを登録して分析が可能。
- 複合語 は俯瞰図上に表示はされるが分析対象とならない。

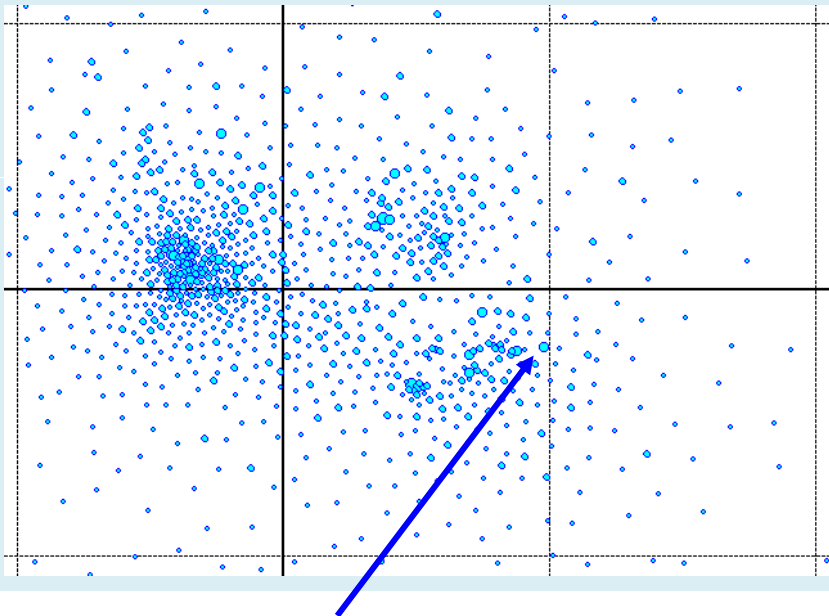
■ DocRadar分析例 1

DocRadarのオフラインでの表示例

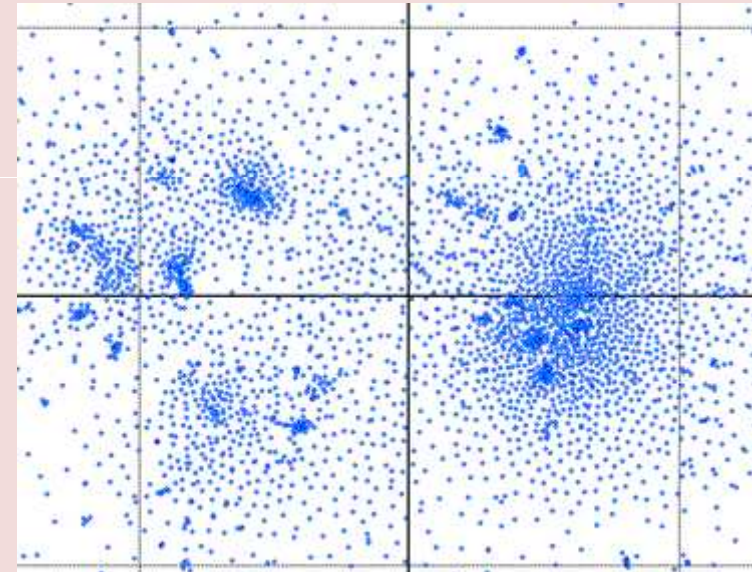


自動運転（日本公報、総文献数4,100件）の全文を分析したもの

- ・ クラスタ表示(初期設定)
類似特許を束にして表示



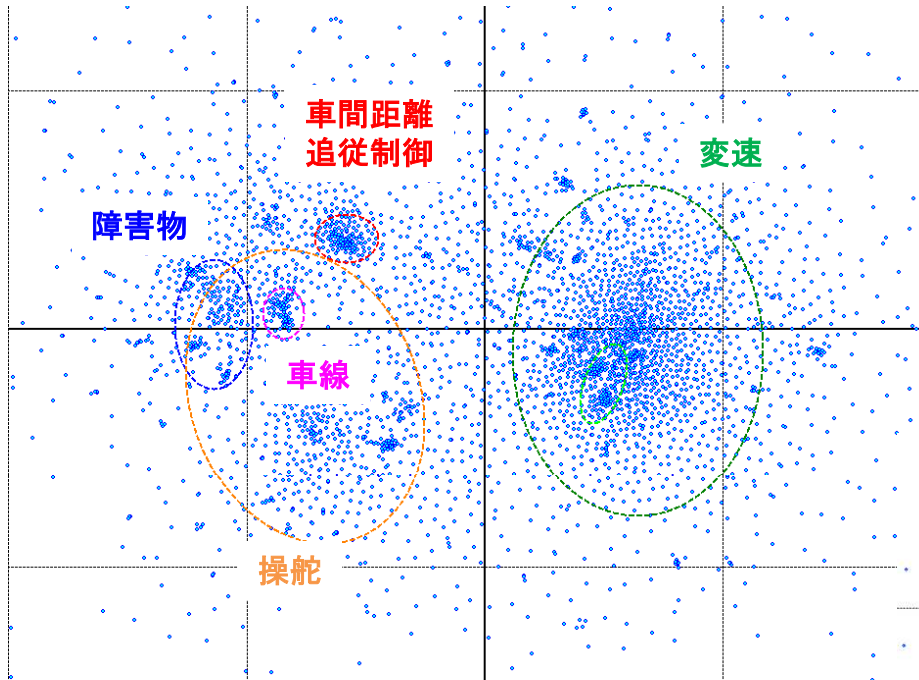
- ・ 一件一葉表示
1公報を1つの点で表示



例えば、このクラスターには類似と判断された28公報が含まれる。
(クラスター化強度は任意に設定可)

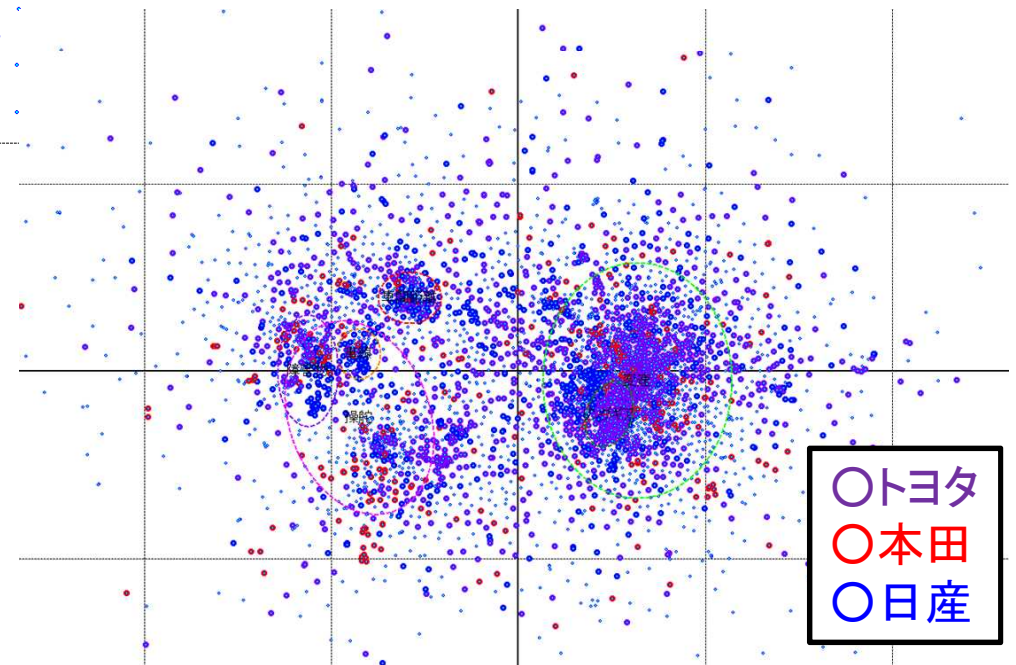
■ DocRadar分析例 2

自動運転の特許に関する特許の分析例



例えば、自社が強み、弱みとしている技術、自社と競合企業の技術開発の方向性等を把握する。

出願人別に表示した例

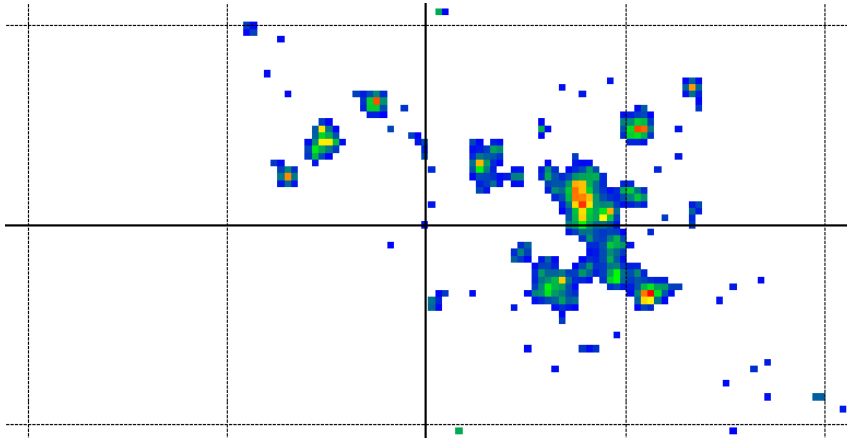


■ DocRadar分析例 3

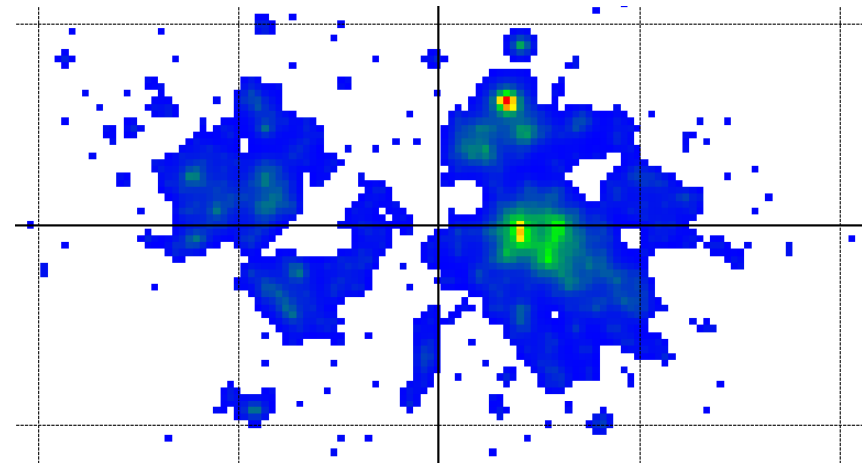
出願年度別に色付けし、開発の経年変化の確認した例



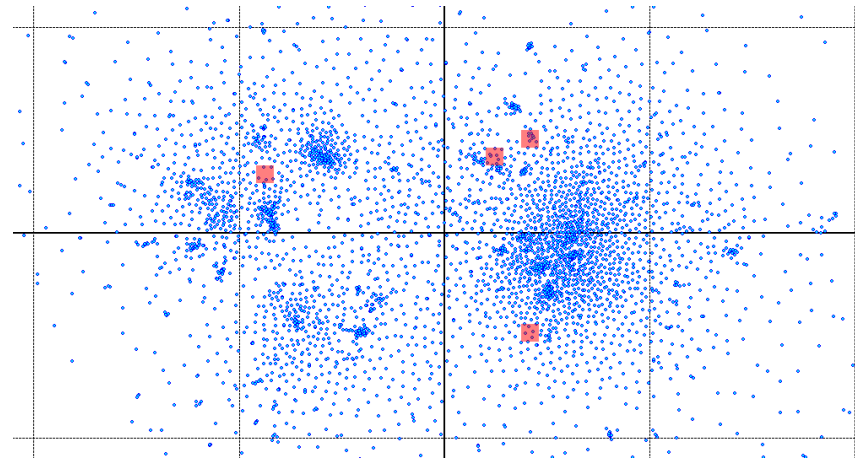
1999~2001



2014~2016



直近5年分の増加率と、出願割合を指定し、近年開発が増加している領域を抽出した例



■ ThemeScapeとDocRadarの比較

	ThemeScape	DocRadar
解析原理(共通)	文章をトピック（特徴語）に分解、各文章の関係を多次元ベクトル空間で計算し、二次元マップ上に表示	
ストップワード （禁止語）	個別設定可能 低頻度ワードの閾値設定可能 数字を含む単語の除外設定可能 □/低頻度ワードを自動除外	個別設定可能（ウエイト=0） 辞書登録可能
シソーラス	設定不可 英単語の語尾変化は自動認識	設定可能 辞書登録可能
フレーズ （複合語）	解析に使用される	抽出、表示されるが解析には使用されない
マップ画面での トピックの表示	自動表示される	自動表示されない
マップ画面での 特許の配置	特許毎に配置	クラスター単位で配置 特許毎の配置も可能
データベース	Derwent Innovationを基本	データベース付属無し
システム構成	ASP/クラウド	ASP/クラウド 分析はオフラインでも可能

【検討詳細】

- 俯瞰解析ソフトの現状と比較
- ThemeScapeの原理、特徴
- DocRadarの原理、特徴
- **DocRadarとThemeScapeの比較**

■ DocRadarとThemeScapeの比較

- 目的

同じ母集合を用いて解析を実行し、その結果を比較することにより2つの俯瞰解析ソフトの違いを明らかにする。

- 結論

DocRadar(TechRadar*)とThemeScapeでは、

**「解析に適する、母集合の技術分野の広さ
(粗さ/細かさ)が異なる」**

ことが、明らかとなった。この結果から、

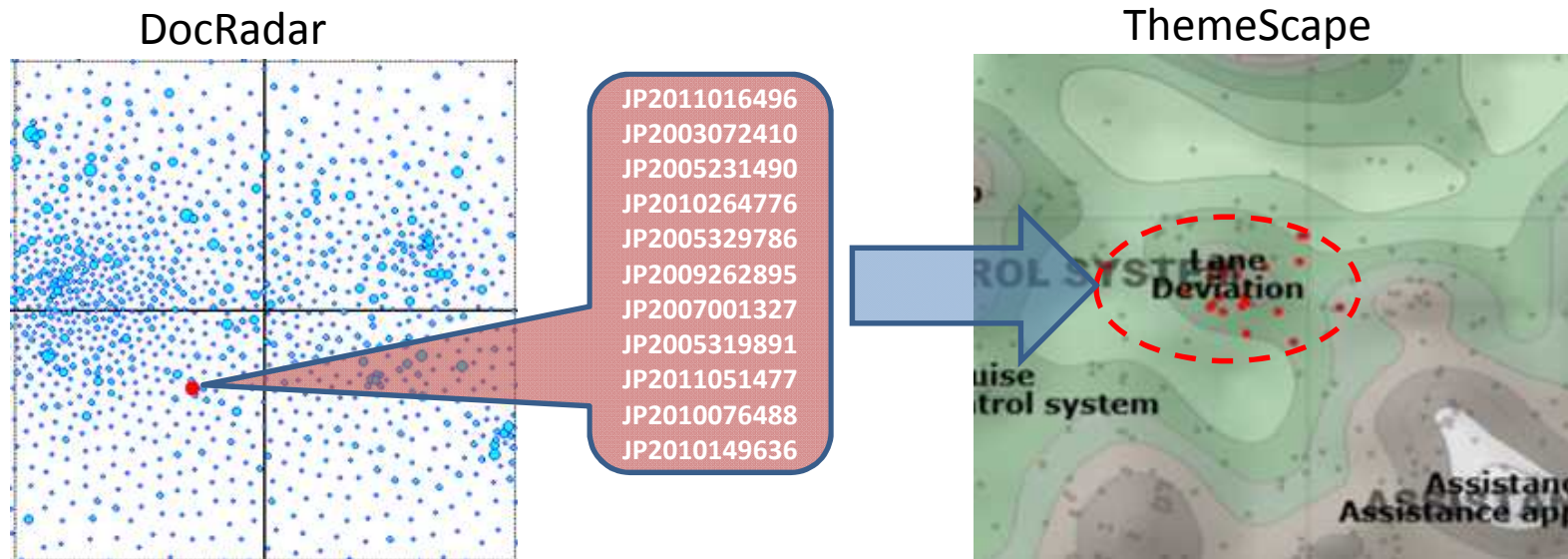
「それぞれのアプリケーションに適した使い方」

を提案する。

*TechRadar：DocRadarを日本語公報解析に特化したシステム

■ 基本方針

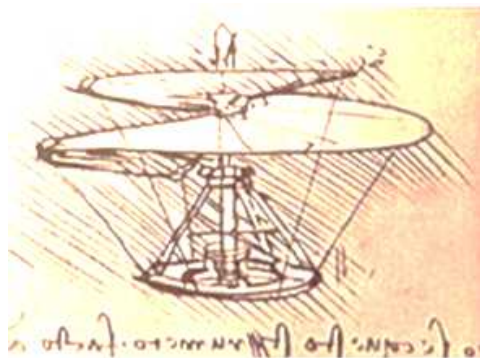
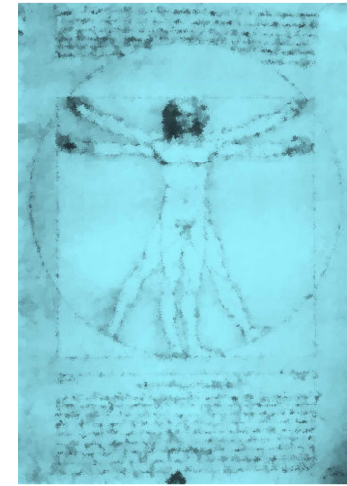
- 1) DocRadarとThemeScapeで同じ文章を解析する。
→ DocRadar にDWPI_抄録を読み込ませて解析。
- 2) DocRadarとThemeScapeで解析する単語を一致させる。
→ ストップワード（禁止語）をできるだけ一致させる。
- 3) DocRadarとThemeScapeで、同一技術と分類される公報が一致するか相関関係を比較する。



DocRadarクラスター内の公報が ThemeScape 上で集中して分布 = 解析に相関性あり

■ 検討

- 1) 母集合の検討
- 2) DWPI_抄録抽出
- 3) ストップワード（禁止語）の検討
- 4) DocRadarとThemeScapeによる解析結果比較
- 5) 考察
- 6) DocRadarとThemeScapeの比較 ～まとめ～



1) 母集合の検討

テーマ 「自動車の自動運転技術」

Fターム 3D241AB01(車両の運動制御 / ・一般車両)

11,322件

→ 2000年以降に限定

4,323件

3D241	駆動装置の関連制御、車両の運動制御											
観点	Fターム											
AA	AA00		AA01		AA03	AA04	AA05	AA06	AA07		AA09	
	駆動装置等の関連制御. 目的		・ 振動防止		・ 耐久性向上	・ ・ クラッチの摩耗, 焼損防止	・ ・ 内燃機関の損傷防止	・ ・ 変速ギヤの破損防止	・ ・ Vベルトの損傷防止		・ 小型軽量化	
			AA11	AA12	AA13		AA15		AA17	AA18	AA19	
			・ 構造簡素化	・ ・ ペダル, レバーの単一化, 兼	・ ・ ・ 身体障害者用		・ 駆動装置間の単なる連動, 自動		・ 公害対策	・ ・ 騒音低減	・ ・ 排気ガス対策	
AB	AB00		AB01	AB02		AB04	AB05		AB07			
	駆動装置等の関連制御. 適用車種		・ 一般車両	・ ・ バス, トラック		・ 農業車両	・ ・ 歩行型		・ 建設, 土木, 産業車両			
AC	AC00		AC01	AC02	AC03	AC04		AC06	AC07	AC08	AC09	AC10
	駆動装置等の関連制御. 構造		・ 推進装置, エンジン	・ ・ ディーゼルエンジン	・ ・ 過給機	・ ・ アクセルペダル, レバー		・ クラッチ	・ ・ 流体圧式	・ ・ トルクコンバータ, 流体継手	・ ・ ・ ロックアップクランチ付きトルクコンバー	・ ・ 電磁式, 電磁粉式

2) DWPI_抄録抽出

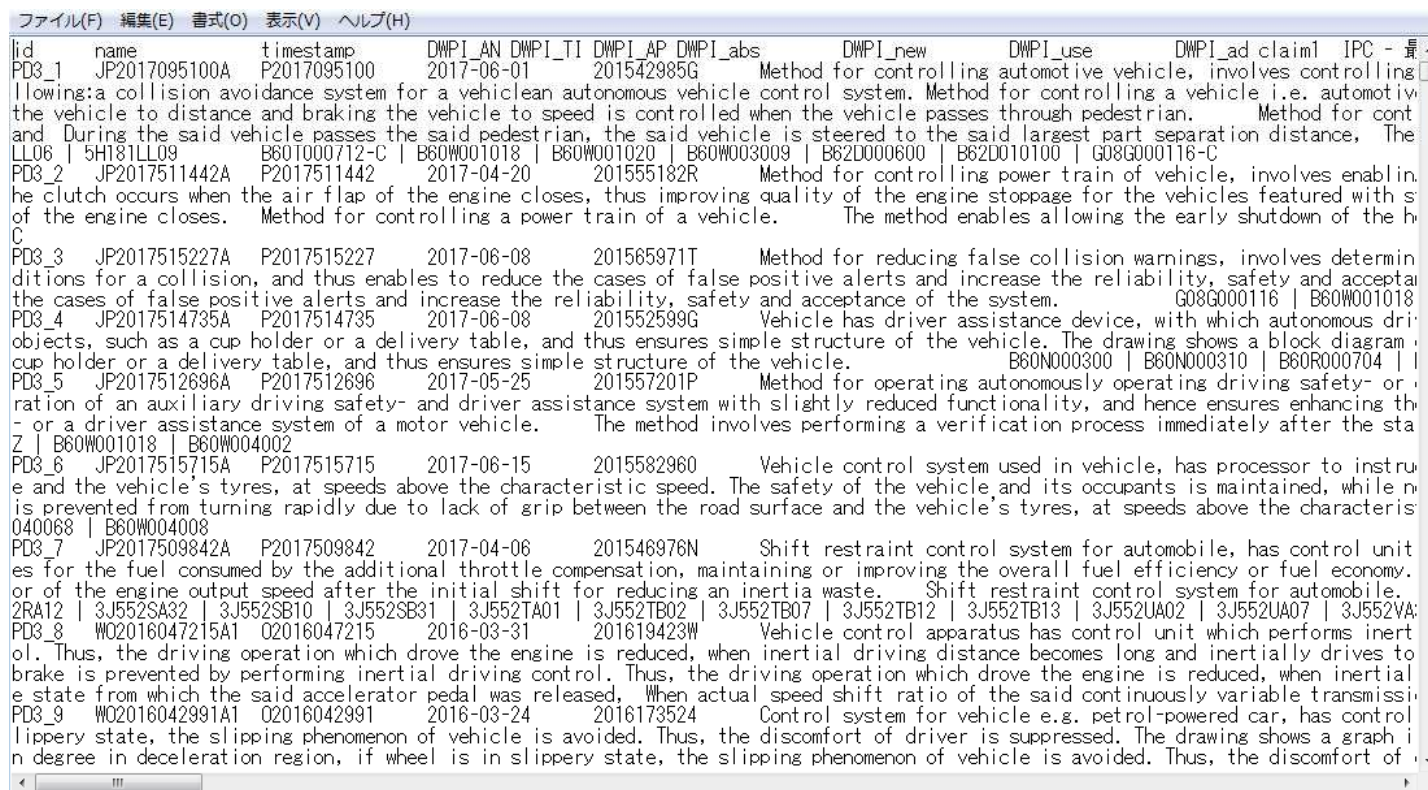
Patent SQUAREでFチームに属する公報番号を抽出



Derwent Innovationで番号検索し、DWPI_抄録を出力



DocRadarの母集合として登録(TXTファイル)



id	name	timestamp	DWPI_AN	DWPI_TI	DWPI_AP	DWPI_abs	DWPI_new	DWPI_use	DWPI_ad	claim1	IPC
PD3_1	JP2017095100A	P2017095100	2017-06-01		201542985G					Method for controlling automotive vehicle, involves controlling following: a collision avoidance system for a vehiclean autonomous vehicle control system. Method for controlling a vehicle i.e. automotiv the vehicle to distance and braking the vehicle to speed is controlled when the vehicle passes through pedestrian. Method for cont and During the said vehicle passes the said pedestrian, the said vehicle is steered to the said largest part separation distance, The LL06 5H181LL09	B60T000712-C B60W001018 B60W001020 B60W003009 B62D000600 B62D010100 G08G000116-C
PD3_2	JP2017511442A	P2017511442	2017-04-20		201555182R					Method for controlling power train of vehicle, involves enablin he clutch occurs when the air flap of the engine closes, thus improving quality of the engine stoppage for the vehicles featured with s of the engine closes. Method for controlling a power train of a vehicle. The method enables allowing the early shutdown of the h C	
PD3_3	JP2017515227A	P2017515227	2017-06-08		201565971T					Method for reducing false collision warnings, involves determin ditions for a collision, and thus enables to reduce the cases of false positive alerts and increase the reliability, safety and accepta the cases of false positive alerts and increase the reliability, safety and acceptance of the system. G08G000116 B60W001018	
PD3_4	JP2017514735A	P2017514735	2017-06-08		201552599G					Vehicle has driver assistance device, with which autonomous dri objects, such as a cup holder or a delivery table, and thus ensures simple structure of the vehicle. The drawing shows a block diagram - cup holder or a delivery table, and thus ensures simple structure of the vehicle. B60N000300 B60N000310 B60R000704	
PD3_5	JP2017512696A	P2017512696	2017-05-25		201557201P					Method for operating autonomously operating driving safety- or ration of an auxilliary driving safety- and driver assistance system with slightly reduced functionality, and hence ensures enhancing th - or a driver assistance system of a motor vehicle. The method involves performing a verification process immediately after the sta Z B60W001018 B60W004002	
PD3_6	JP2017515715A	P2017515715	2017-06-15		2015582960					Vehicle control system used in vehicle, has processor to instru e and the vehicle's tyres, at speeds above the characteristic speed. The safety of the vehicle and its occupants is maintained, while n is prevented from turning rapidly due to lack of grip between the road surface and the vehicle's tyres, at speeds above the characteris 040068 B60W004008	
PD3_7	JP2017509842A	P2017509842	2017-04-06		201546976N					Shift restraint control system for automobile, has control unit es for the fuel consumed by the additional throttle compensation, maintaining or improving the overall fuel efficiency or fuel economy. or of the engine output speed after the initial shift for reducing an inertia waste. Shift restraint control system for automobile. 2RA12 3J552SA32 3J552SB10 3J552SB31 3J552TA01 3J552TB02 3J552TB07 3J552TB12 3J552TB13 3J552UA02 3J552UA07 3J552VA	
PD3_8	WO2016047215A1	O2016047215	2016-03-31		201619423W					Vehicle control apparatus has control unit which performs inert ol. Thus, the driving operation which drove the engine is reduced, when inertial driving distance becomes long and inertially drives to brake is prevented by performing inertial driving control. Thus, the driving operation which drove the engine is reduced, when inertial e state from which the said accelerator pedal was released. When actual speed shift ratio of the said continuously variable transmissi	
PD3_9	WO2016042991A1	O2016042991	2016-03-24		2016173524					Control system for vehicle e.g. petrol-powered car, has control lippery state, the slipping phenomenon of vehicle is avoided. Thus, the discomfort of driver is suppressed. The drawing shows a graph i n degree in deceleration region, if wheel is in slippery state, the slipping phenomenon of vehicle is avoided. Thus, the discomfort of	

3) ストップワード（禁止語）の検討

活動当初は、解析に使用する単語を一致させる予定だったが、DocRadarとThemeScapeで解析対象が大きく異なり、一致させることは困難なことが判明した。

	DocRadar	ThemeScape
解析単語数	数万語	数百語
複合語	非解析対象	解析対象
語尾変化	一部同一単語 として対応	同一単語扱い

⇒ **ストップワード（禁止語）** はそれぞれ指定できるため、これをできるだけ合わせる方向で検討した。

DocRadar : 特徴語(解析に使用した単語)から、
使ってはいけない単語を指定。 → 禁止語

ThemeScape : 解析に使用しない単語を設定。 → ストップワード

ストップワード（禁止語）の設定ルール

ThemeScapeの設定

- 特徴語で使用されている単語はストップワードから外す。
- アルファベット2文字以下の明らかに意味のない記号はストップワードに追加。

DocRadarの設定

- 数字の付いた特徴語は禁止語とする。
- アルファベット2文字以下の明らかに意味のない記号は禁止語とする。



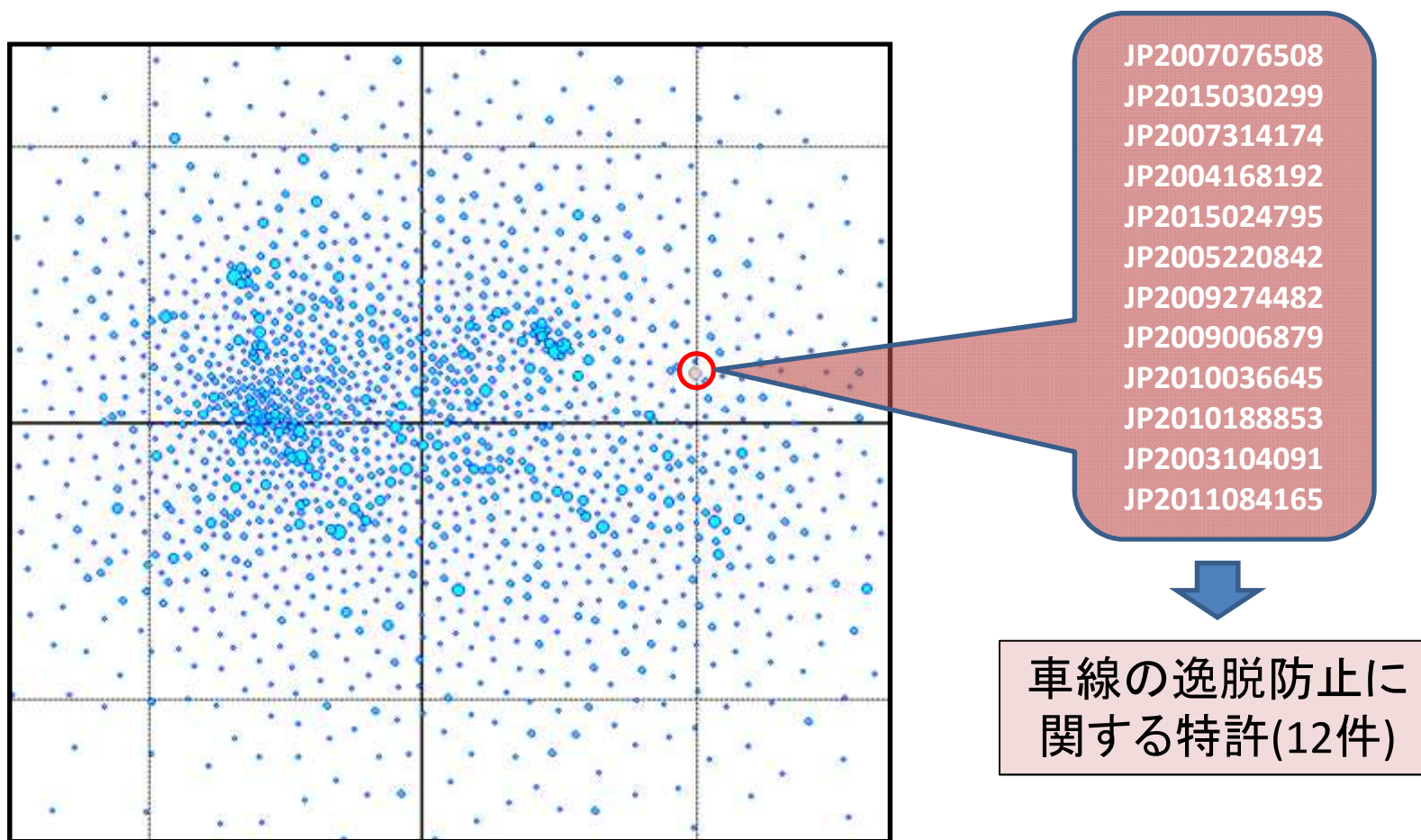
特徴語とストップワードの一覧表を作り、
DocRadar、ThemeScape にそれぞれ設定する。



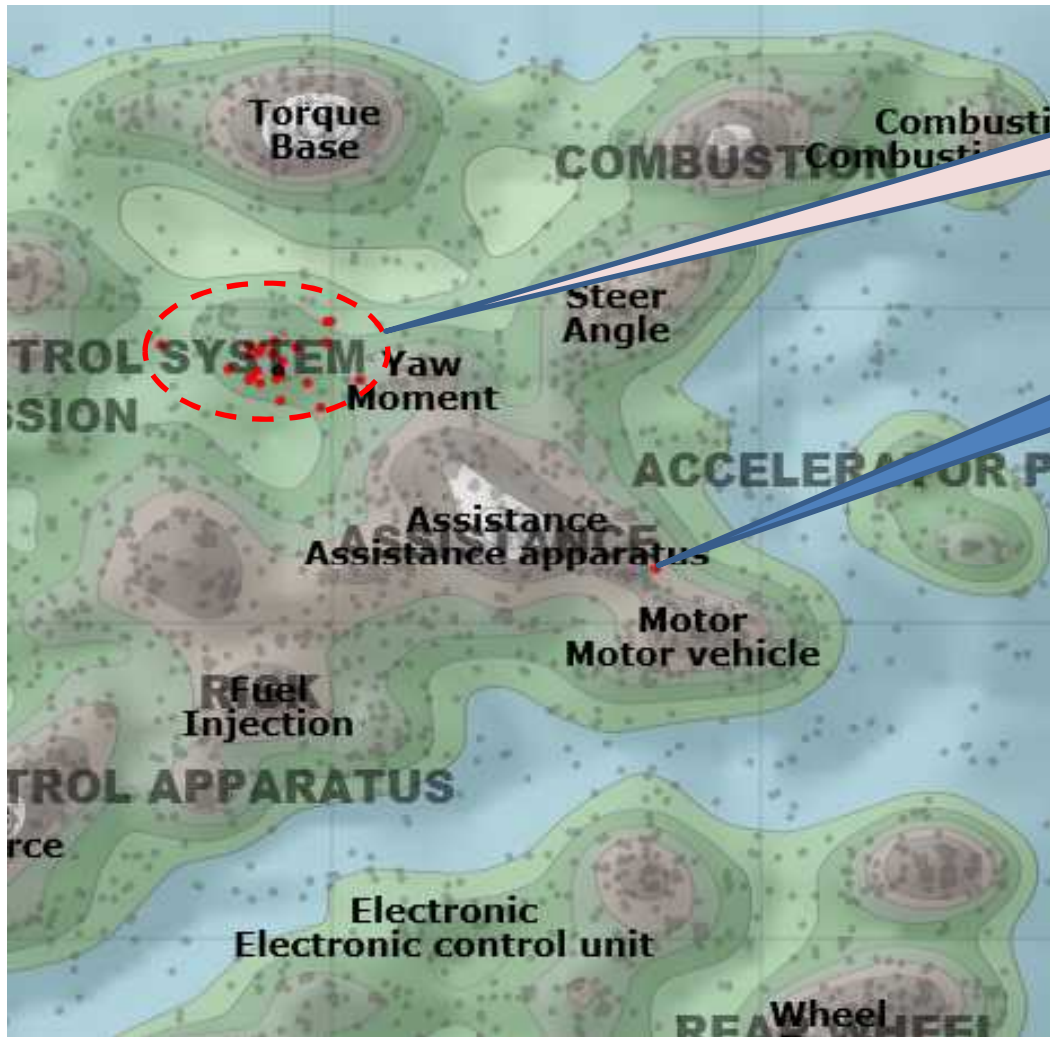
解析結果を比較するための準備完了!!



4) DocRadarとThemeScapeによる解析結果比較 ～解析結果比較1～



DocRadarによる俯瞰図(クラスター弱)



ほぼ一ヶ所に集中
(11/12)

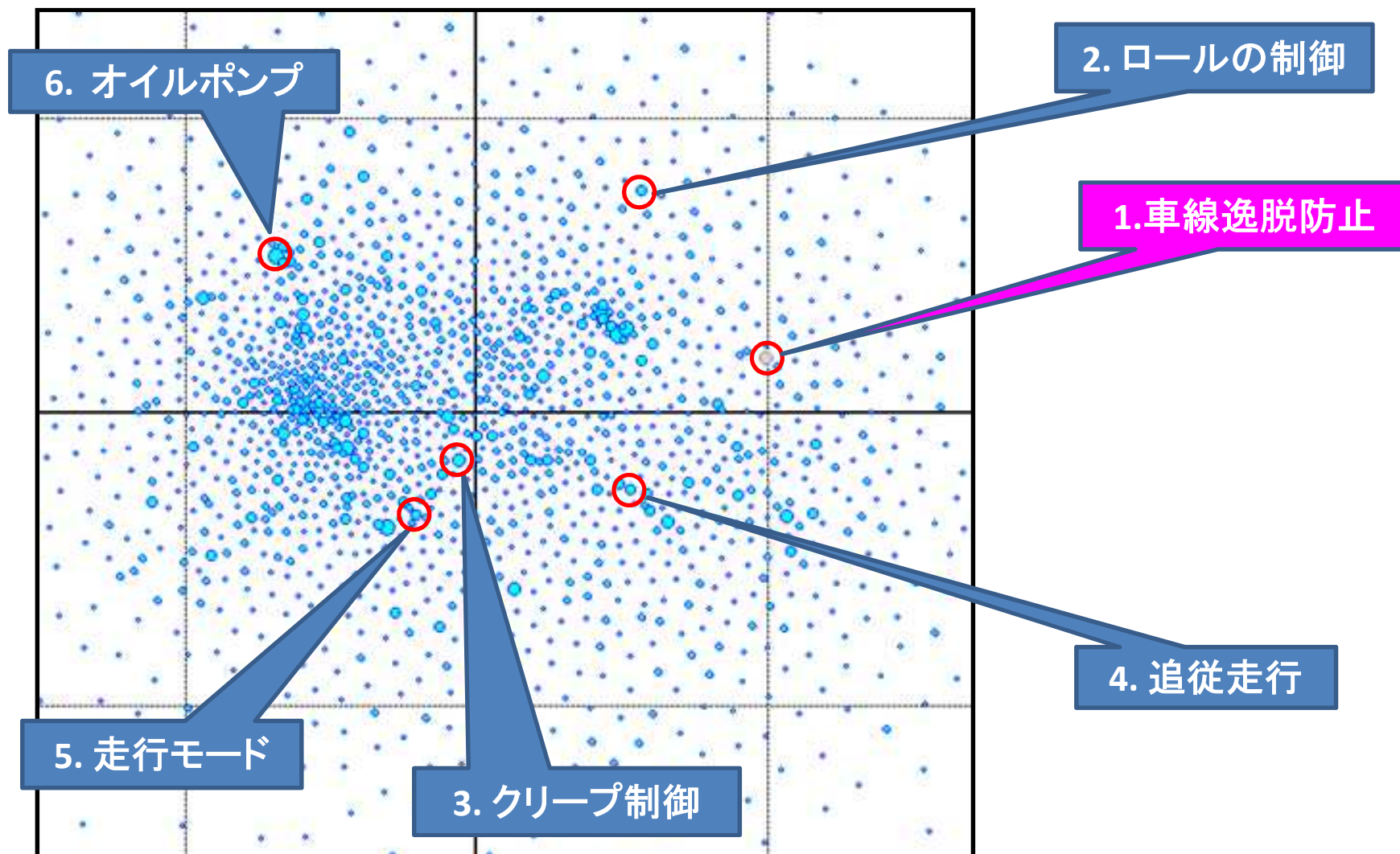
車線維持制御を、四輪
のトルク制御で行う内容

相関性は
非常に高い



ThemeScapeによる俯瞰図

～解析結果比較2～



DocRadarによる俯瞰図(クラスター弱)



3. クリープ制御

1. 車線逸脱防止

6. オイルポンプ

2. ロールの制御

4. 追従走行

5. 走行モード

相関のある集合
と相関の無い
集合が混在



ThemeScape俯瞰図への投影

DocRadar でクラスターに分類された特許番号をThemeScape
上で検索すると、**相関のある集合と相関の無い集合が混在。**
(傾向としては、相関が無い集合の方が多い)

何故か?

- 特徴語、ストップワードの調整 ⇒ 相関性向上せず。
- 相関性の高い集合と、相関性の低い集合の特許の内容比較
⇒ 明確な違いは見られず。

問題点整理

- 1) DocRadarの解析精度?
- 2) ThemeScapeの解析精度?
- 3) DWPI_抄録の記載内容?

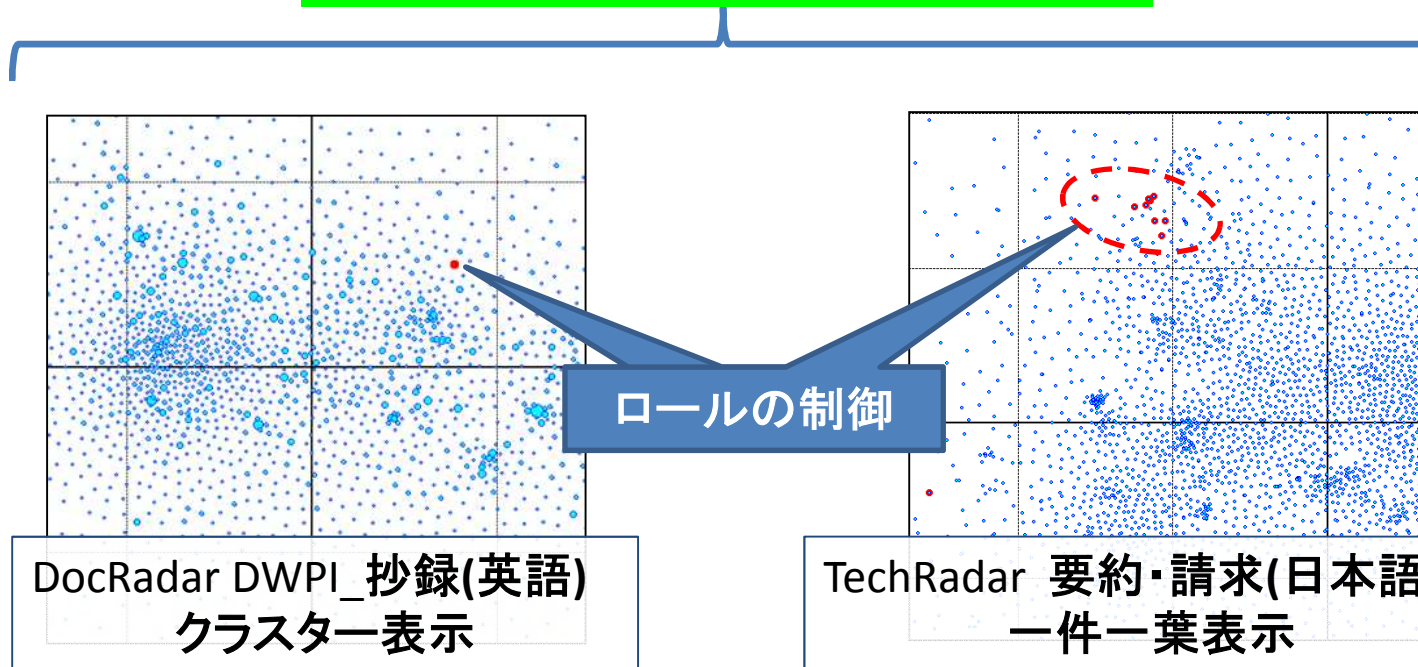


～解析結果比較3～

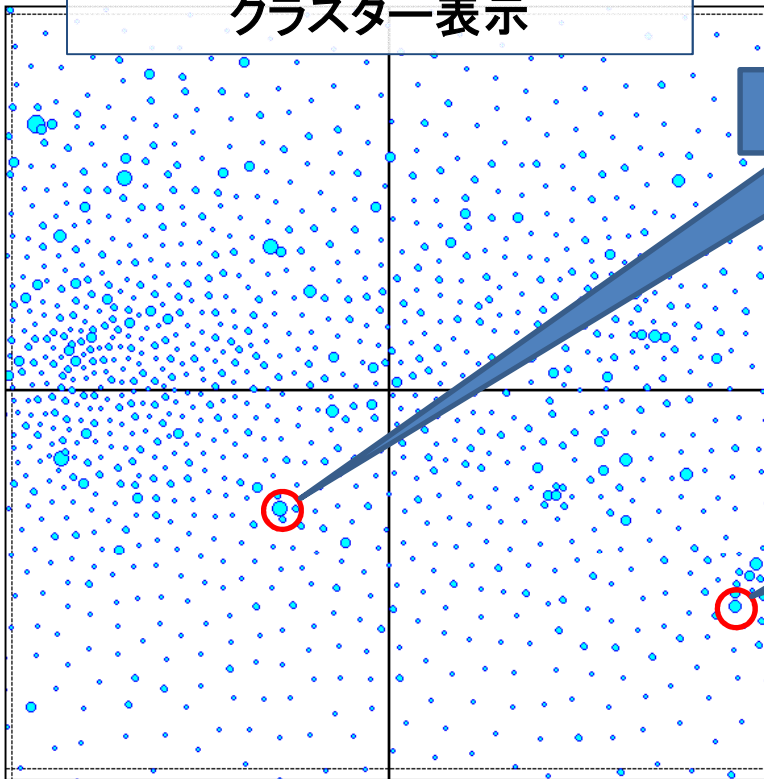
- 状況打開のため、DocRadar、TechRadar、ThemeScapeで様々なチャートを作成し、それぞれの相関性を比較した。

例) FI分類の分布、DocRadar一件一葉表示、TechRadar全文解析、TechRadar要約請求項解析 等

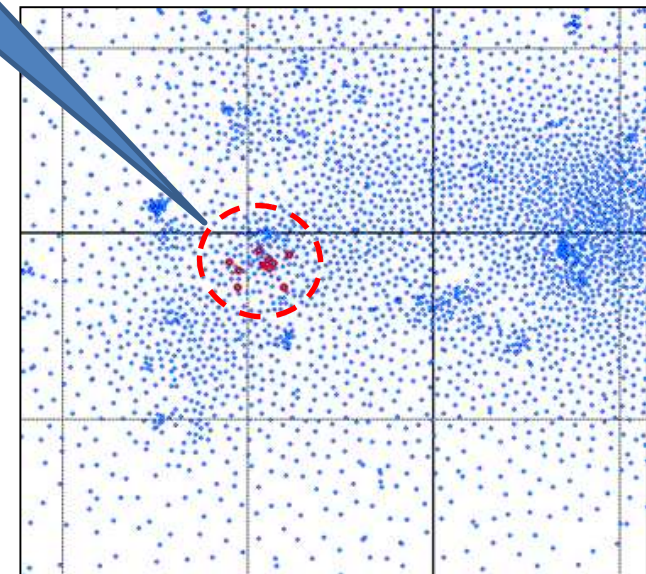
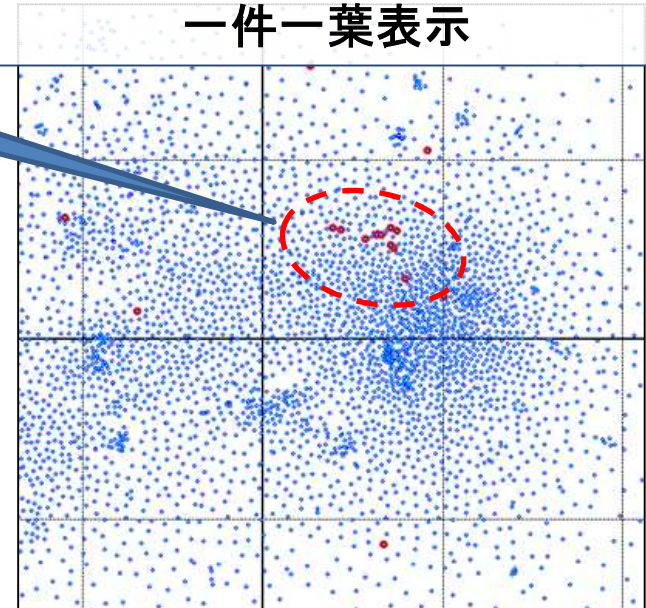
相関性のある組み合わせを発見



DocRadar DWPI_抄録(英語)
クラスター表示



TechRadar 要約・請求(日本語)
一件一葉表示



- 1) DocRadar、TechRadar解析の再現性は高い。
- 2) DWPI_抄録は要約・請求項の内容を正確に反映している。

5) 考察

● 問題点

- 1) DocRadarの解析精度?
- 2) ThemeScapeの解析精度?
- 3) DWPI_抄録の記載内容?



● 判ったこと

- 1) DocRadar、TechRadar解析の再現性は高い。
- 2) DWPI_抄録は要約・請求項の内容を正確に反映している。

● 一般的な知見

ThemeScapeで技術分野を分離した例は多い。



● 推測

今回の「自動運転」の集合は、ThemeScapeで解析するには
「細か過ぎる」
のではないか?



● 検証

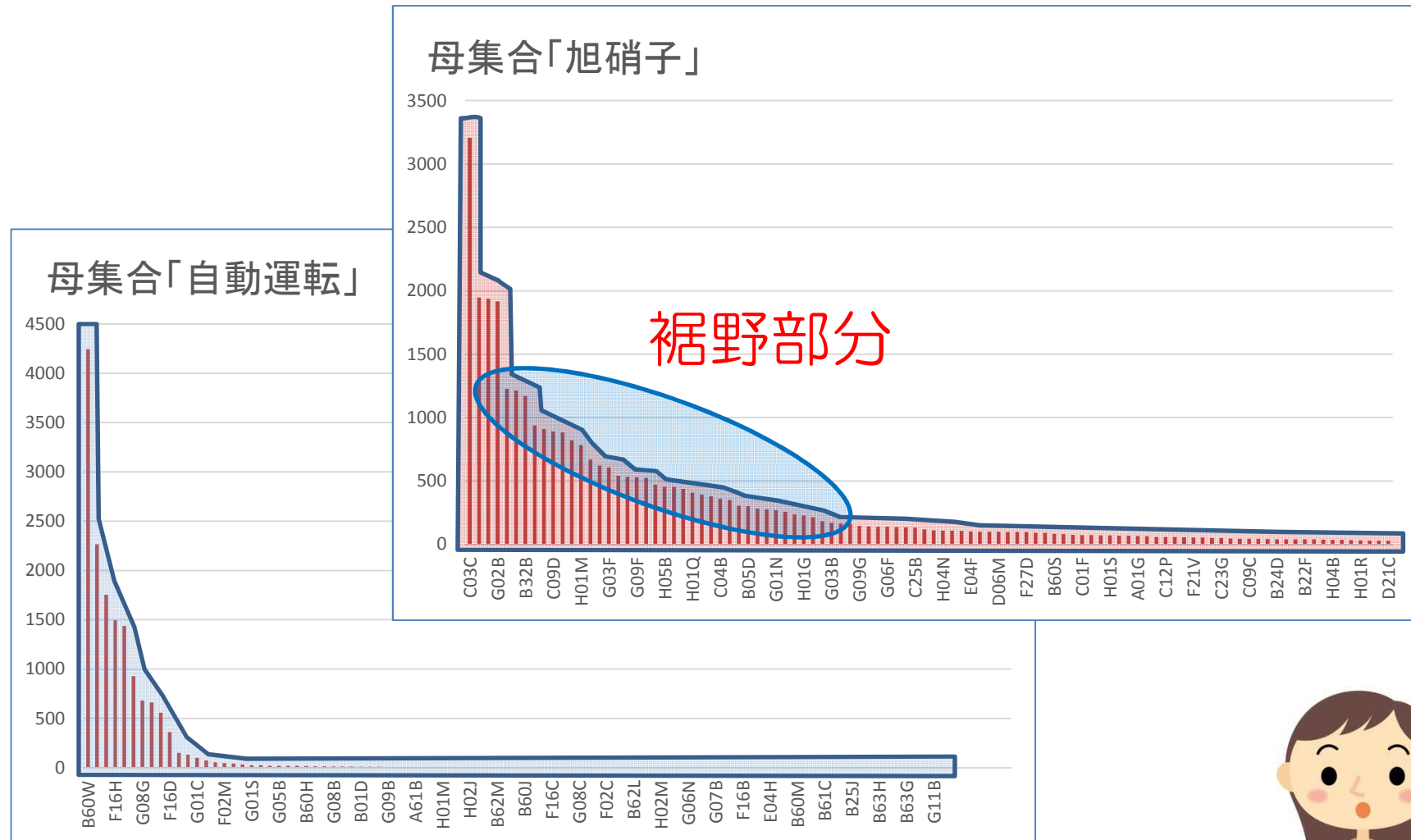
もう少し広い技術分野の集合で同様な比較を行う。



～考察の検証1～

- 母集合 出願人 「旭硝子」
1997年～2017年出願に限定
- TechRadarで検索 **12,212**件 → 要約・請求項で解析
- Patent SQUAREで公報番号を抽出
↓
Derwent Innovationで番号検索
ファミリー取得 **12,231**件 → DWPI_抄録で解析
- ストップワード、禁止語は調整せず

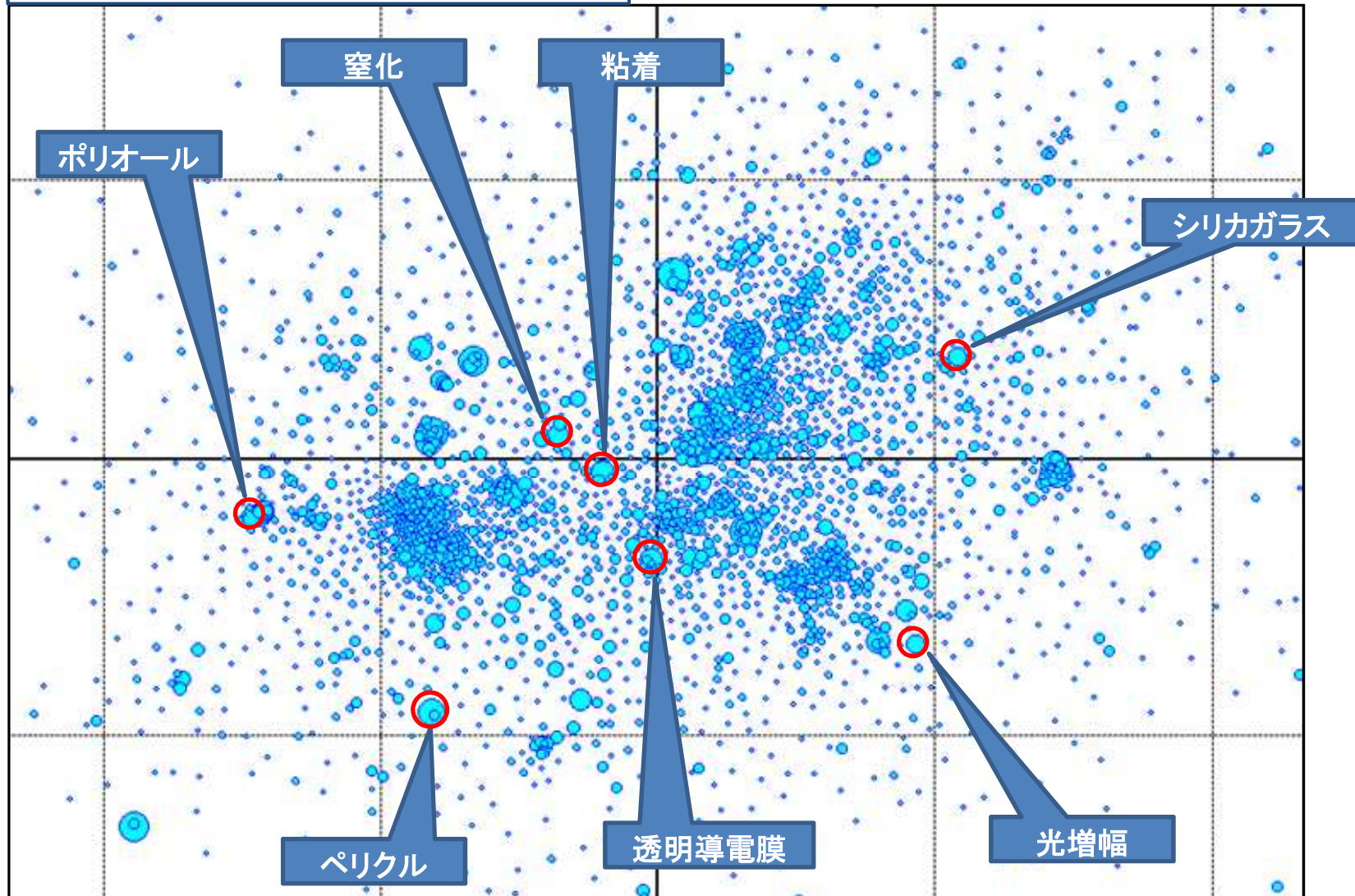
■ 技術分野の広さ（検証1のIPC分布比較）



- 「自動運転」の集合のIPCランキングマップはシャープな形状をしているのに対し、「旭硝子」は裾野が広がっており相対的に技術分野が広いことが伺える。



7個のクラスターについて検証



旭硝子のTechRadarによる解析結果(クラスター弱)



ThemeScape俯瞰図の相関性

～考察の検証2～

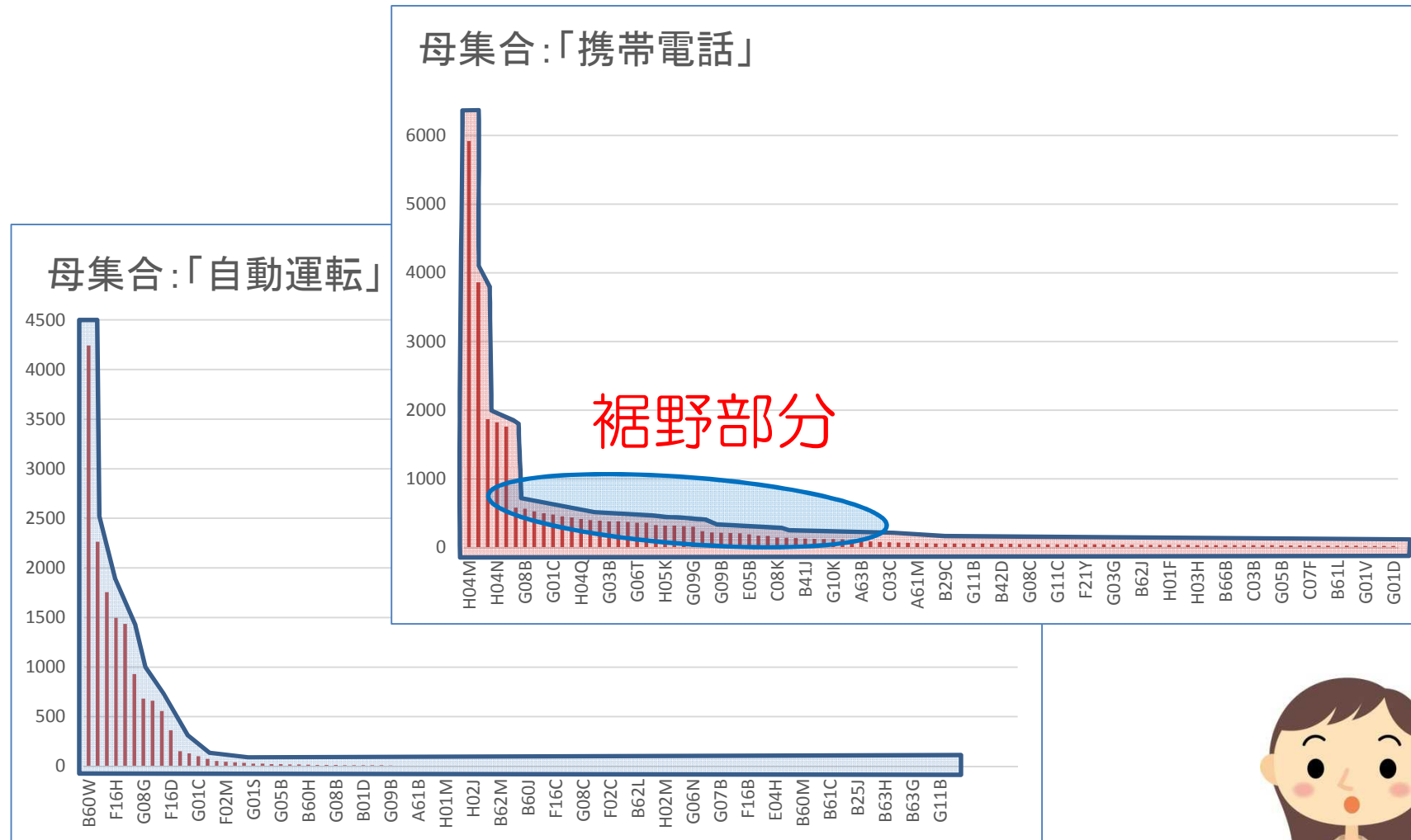
- 母集合 「携帯電話」
- PatentSQUAREで検索 17,216件

行	項目	検索式
1	四法	P
2	出願日	20080101～
3	名称、要約、請求項	[?携帯?、?モバイル?、?スマート?*?電話?、?フォン?]W10
4	理論式	1 and 2 and 3

※上記でヒットしたリストのうち、日本特許に限定(PCT日本語出願を除去)

- TechRadarで番号検索 15,997件
→ 要約・請求項で解析
- Derwent Innovationで番号検索 15,399件
→ DWPI_抄録で解析
- ストップワード、禁止語は調整せず

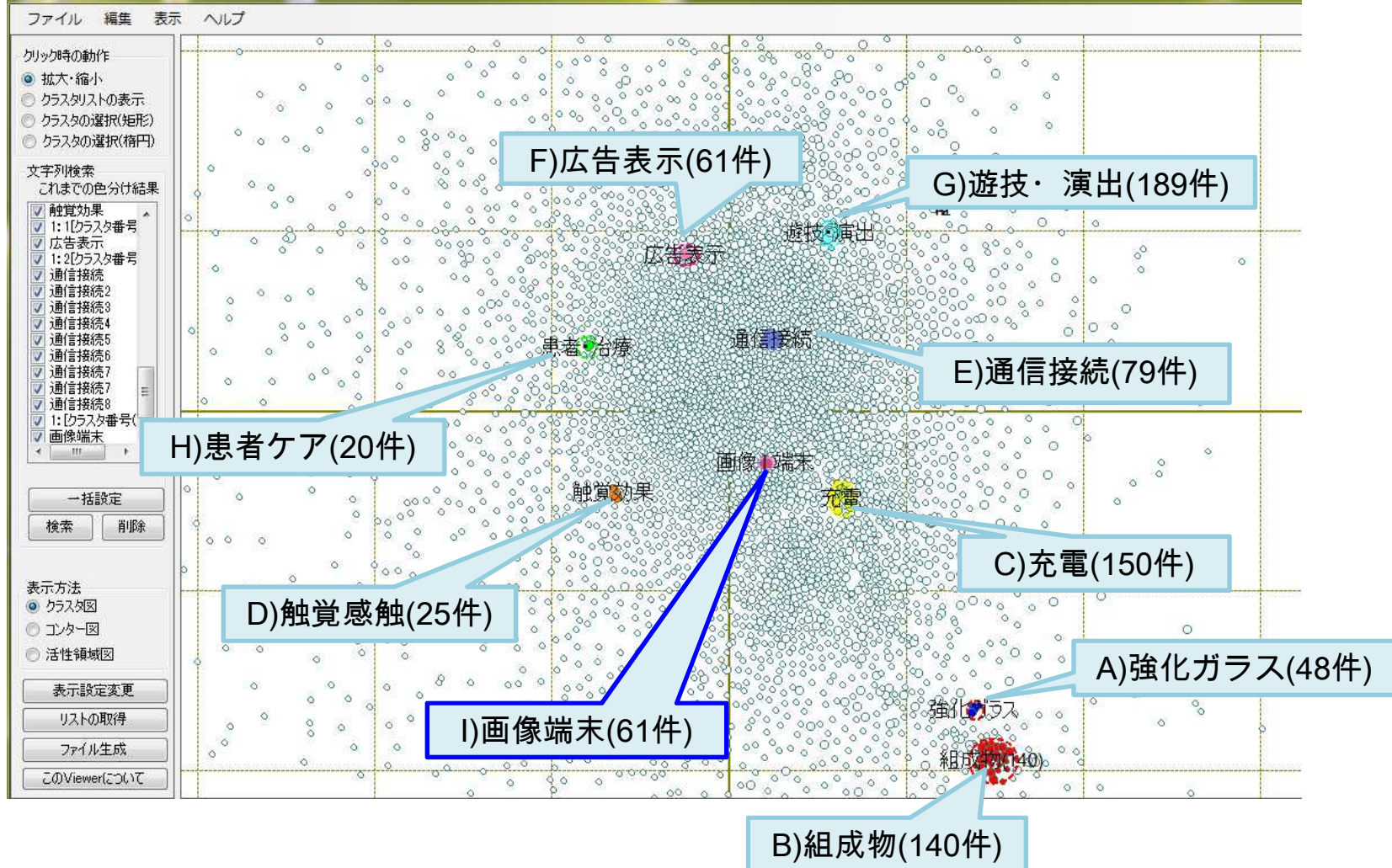
■ 技術分野の広さ（検証2のIPC分布比較）



- 「自動運転」の集合のIPCランキングマップはシャープな形状をしているのに対し、「携帯電話」は裾野が広がっており相対的に技術分野が広いことが伺える。



技術分野が同じ複数クラスターの集合(9個)について検証



携帯電話のTechRadarによる解析結果(クラスター弱)



ThemeScape俯瞰図の相関性

6) DocRadarとThemeScapeの比較 ～まとめ～

- TechRadar(要約・請求項(日本語))解析と ThemeScape(DWPI_抄録(英語))解析の相関性を調べた結果、

(検証1)	一ヶ所に集中して分布	4テーマ	相関有
	二ヶ所に集中して分布	2テーマ	相関有
	分散した	1テーマ	相関無
(検証2)	一ヶ所に集中して分布	8テーマ	相関有
	分散した	1テーマ	相関無



→ **広い技術分野**を含む母集合では、TechRadarとThemeScapeの**解析結果に相関性が見られた**。

< 結論 >

DocRadar(TechRadar)とThemeScapeは、

- **同じ技術内容の公報を、近い位置に集める能力は同じ。解析に適する”技術分野の広さ”に違いがある。**

〈提案〉 それぞれのアプリケーションに適した使い方

● ThemeScape

各種の分析ツールが充実していて操作も容易なため、技術的に土地勘のない分野において、技術の違いが明確である程度広い技術分野を含む集合の総括的な解析に適している。

● DocRadar (TechRadar)

広い技術分野の解析もできるが、単語調整とシソーラス調整を細かく行うことで、細かい技術の違いが分離できるため、よく知っている技術分野の詳細な解析に適している。

本年4月に「集合を代表するキーワードを抽出する機能を追加した。」という情報あり。



【結果とまとめ】

● ThemeScape の特徴

- ThemeScape の概要とマップ生成原理を調査した。
- 単語調整の特徴として、①シソーラスはない、②ストップワードを設定可能、③単語とフレーズは別途カウントされることを確認した。

● DocRadar の特徴

- 形態素解析、特徴語抽出、類似度算出から可視化といった、大まかな解析処理フローにはThemeScape と大きな差はない。
- ①禁止語・シソーラスを設定可能、②複合語は解析には使用されない、③クラスター単位で解析可能、④特徴語のマップ上への自動表示は不可であることを確認した。

● ThemeScape と DocRadar の比較

- 同じ母集合を用いて、ストップワード（禁止語）をできるだけ合わせる方向で、2つの解析ツールを比較した。
- 2つの解析ツールは、同じ技術内容の公報を、近い位置に集める能力は同じであるが、解析に適する”技術分野の広がり”に違いがあることが判明した。

【謝辞】

- WGの活動に多大なご協力を賜り、また今回の発表にご快諾いただいた、

Clarivate Analytics社様
VALUENEX社様

この場を借りて感謝申し上げます。

【参考、引用文献など】

- 2015年度プラスドック協議会Bグループ発表資料
- 2016年度プラスドック協議会Dグループ発表資料
- 企業内情報システムについて ―ビッグデータ解析を踏まえて―
知財管理 Vol.67 No9(2017), P1401~P1416

ご清聴ありがとうございました。



【ベンダーへの要望】

● Clarivate Analytics社様 (ThemeScape)

- ドキュメントビューアー内のトピックリストをコピー・利用できるようにしてほしい。
- シソーラス調整ができるようにしてほしい。
- 低頻度ワードの閾値調整時に、閾値と除外ワードの関係を確認できるようにしてほしい。

● VALUENEX社様 (DocRader、TechRadar)

- ダッシュボードは高スペックのパソコンを要求するが、使い勝手が悪いので何とかしてほしい。
- 全体に広く分布する特徴語を自動で抽出できる様にできないか？
- データ登録、viewerがUTF-8形式にしか対応していないので、コード形式を意識せずに使えるようにしてほしい。
- Viewerでの複数の検索が一度に出来るようにしてほしい。