

# KHCoderによる 中国特許の分析可能性評価

第13回情報プロフェッショナルシンポジウム  
2016年12月2日

松下電器研究開発（蘇州）有限公司

○袁 方  
小林 義典  
安藤 俊幸

花王株式会社

## ■背景

---

### ■① KHCoderの評価

- 評価の基準
  - 分かち書きの比較
  - 可視化能力の比較
  - KHCoderの長所と注意点
  - 手動分析との比較
- 

### ■② KHCoderの分析事例

- 特定会社に対する出願の分析
- 

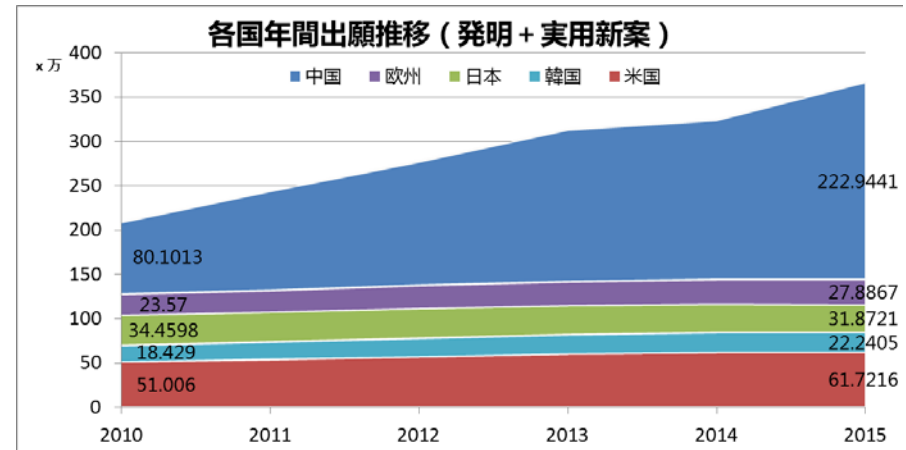
### ■③ ツールの組み合わせ分析

- ツールの組み合わせ分析 \_ RostNat
  - ツールの組み合わせ分析 \_ incoPat
- 

## ■まとめ

## < 背景 >

中国毎年の出願件数も年々増え、2015年の発明・実用新案の合計出願数は222.9万件になっている。

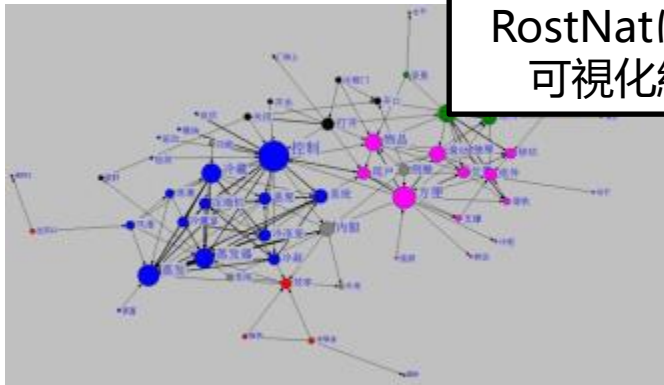


※ 各国特許局公表データで整理

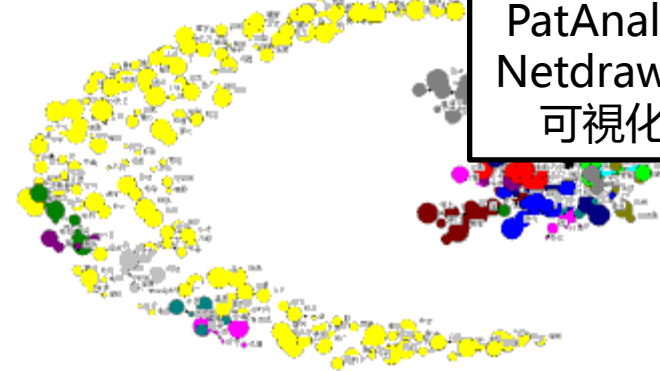
## < 現状 >

中国の出願分析をする際、商用の特許分析ツールのほとんどは『要約の英訳』等に対して『英文のテキストマイニング』を実施し、単語の抽出、分析、及びに可視化を行っている。中国語原文で分析可能なツールは少なく、分かち書きができて、可視化の種類が少ない、等の課題がある

RostNatによる  
可視化結果



PatAnalyzer +  
Netdrawによる  
可視化結果



## 「課題」

- 誤訳
- データ量制限
- 全文分析難あり（訳文入手面）
- 分析&可視化の種類が少ない
- 萌芽KWの抽出に課題

## KHCoder

- 日本語/英語でのテキストマイニングでの実績
- 分析手法が多い（多次元尺度分析、対応分析、階層クラスタ分析、等）
- 操作性（操作簡便性+柔軟性（カスタマイズ一部可能））
- Ver. 3.Alpha.1版～中国語データ分析に対応  
⇒ KHCoderによる『中国語特許データの分析可能性を評価』の必要性

筆者が過去に評価し、十分利用に耐えると判断した中文対応テキストマイニングツールである『RostNatとPatAnalyzer』と比較し、複数の視点でKHCoderの中国語対応機能を検証した。

# ① KHCoderの評価

ソフトの操作性、データ扱いの柔軟性、分かち書きの精度、処理速度と可視化の能力の面で、3つのソフトで比較し、評価を行った

ツールの  
使いやすさ

データの  
取り扱い

	操作性	柔軟性	分かち書き 精度	処理時間	可視化 種類
<b>PatAnalyzer</b> Ver:1.2.17	△	◎	○	△	△
<b>KHCoder</b> Ver:3.Alpha.01a	◎	△	◎	△～×	◎
<b>RostNat</b> Ver:2.0	◎	○	◎	◎	△

## 【操作性】

PatAnalyzer：分かち書き機能のみ、描画にはNetDrawが必要

⇒ワンクリックで実行できるRostNatと、KHCoderより扱いにくい

## 【柔軟性】

PatAnalyzer/RostNat：元データを読み込み、分かち書き等の中間処理後の『中間データ』を編集可能

KHCoder：データ扱いについて自由度が低い（制限大）

# 分かち書きの比較

## 速度

	PatAnalyzer	KHCoder	RostNat
257件要約の 分かち書き	1分31秒	58秒	11秒

同じデータ（空調分野A社、257件の中国特許出願の要約部分）に対して、KHCoder、PatAnalyzerとRostNatとの分かち書きの比較を行った。

## 柔軟性

	PatAnalyzer	KHCoder	RostNat
フィルター	○	○	○
強制接合	×	○	×

特許専門用語排除機能：各ツールとも保有

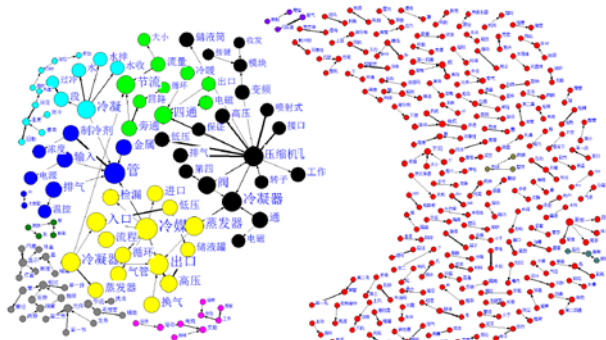
特定の技術用語抽出するための  
強制接合機能はKHCoderのみ保有

## 分かち書き

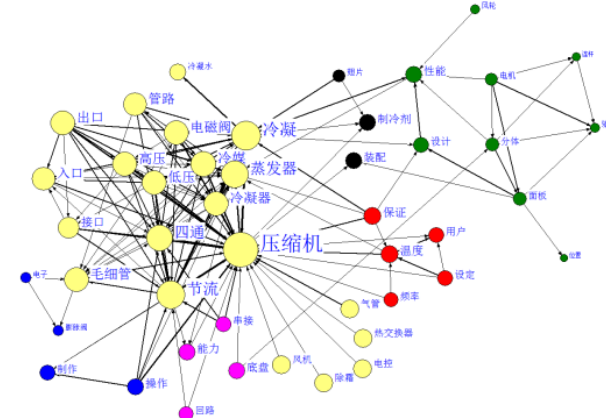
PatAnalyzer		KHCoder		RostNat	
閥	261	圧縮機	245	圧縮機	235
圧縮機	247	閥	185	节流	167
节流	182	节流	157	面板	140
面板	166	面板	155	翅片	111
冷凝	166	管	140	四通	110
板	149	蒸发器	121	蒸发器	109
管	138	冷媒	107	冷媒	107
蒸发器	121	板	106	冷凝器	81
四通	110	翅片	105	电机	79
冷媒	108	电机	97	制冷剂	78
翅片	107	冷凝器	82	温度	75
电机	97	管路	77	管路	66

抽出されたコンセプト語の上位順から比較された結果、KHCoderの分かち書きは他の2つのツールとほぼ同等な精度を有すると判断

# 可視化能力の比較

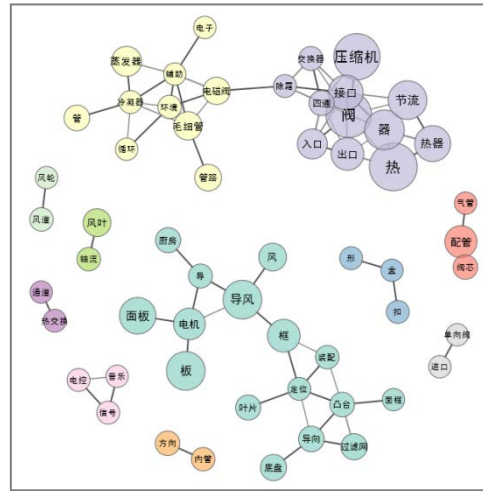


PatAnalyzer+Ucinet

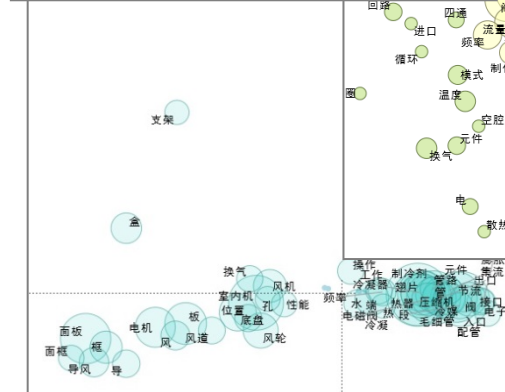


RostNat

## 従来の可視化ツール

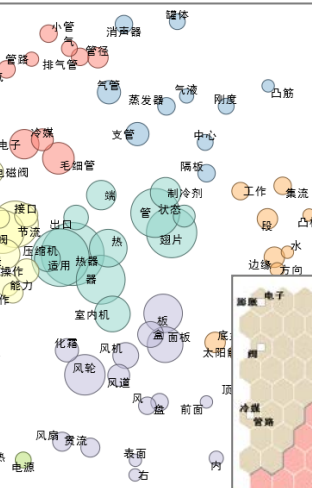
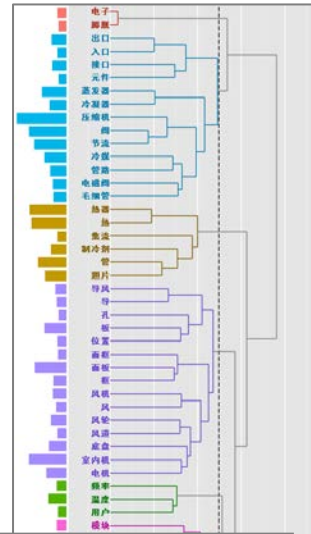


ネットワーク分析

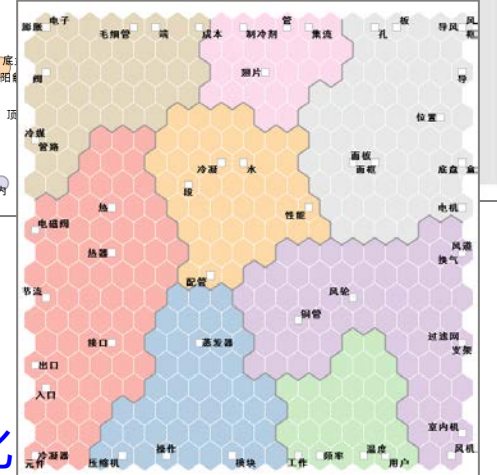


対応分析

階層  
クラスター



多次元尺度



自己組織

## KHCoderの可視化

KHCoderの可視化種類が豊富だが、描画数200を超えると極端に遅くなる

		PatAnalyzer	KHCoder	RostNat
描画数	100	2秒	7秒	2秒
	200	7秒	40秒	7秒
	400	22秒	40分22秒	22秒



## 長所：

- 分かち書き能力は他の有効なツールとほぼ同等の精度
  - 特定の技術用語抽出するための強制接合機能を有するため、分かち書きの精度を調整できる
  - 可視化の種類が多く、多視点で分析可能
- 

## 注意点：

- ノードと該当特許の特定ができない
- 外部変数を使用する際は、元データの中にある内容しか使えないため、外部変数を編集した場合は、分析を再度最初から実施する必要がある
- 語の取捨選択の「ファイルから読み込み」機能：中国語分析では利用不可（txtファイルしか読み込めないため、中国語txtは文字化けする）

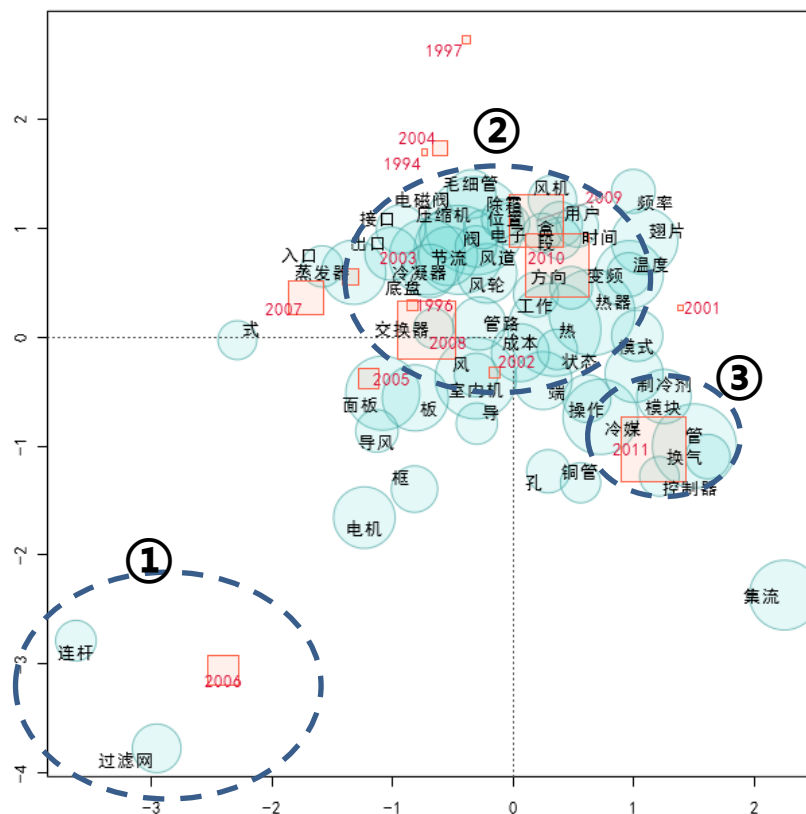
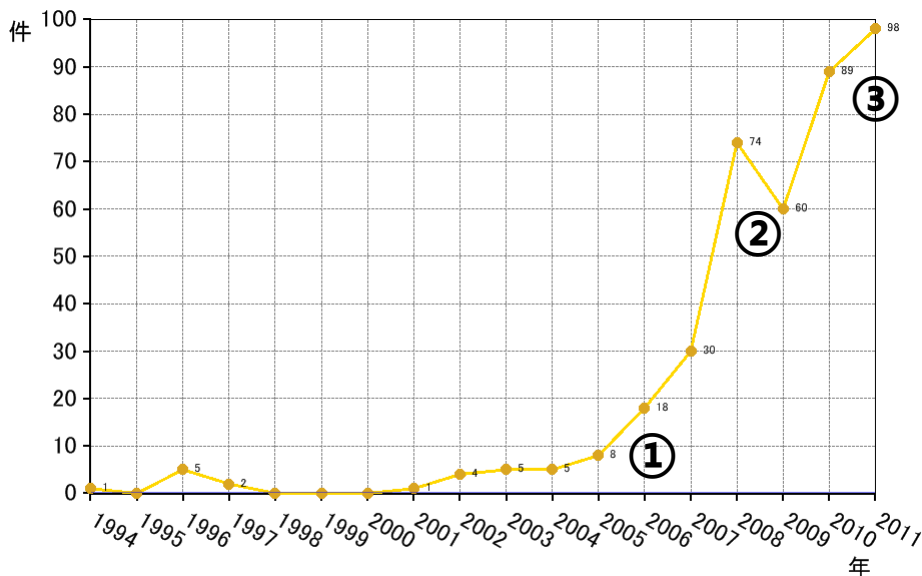
## ②KHCoderの分析事例

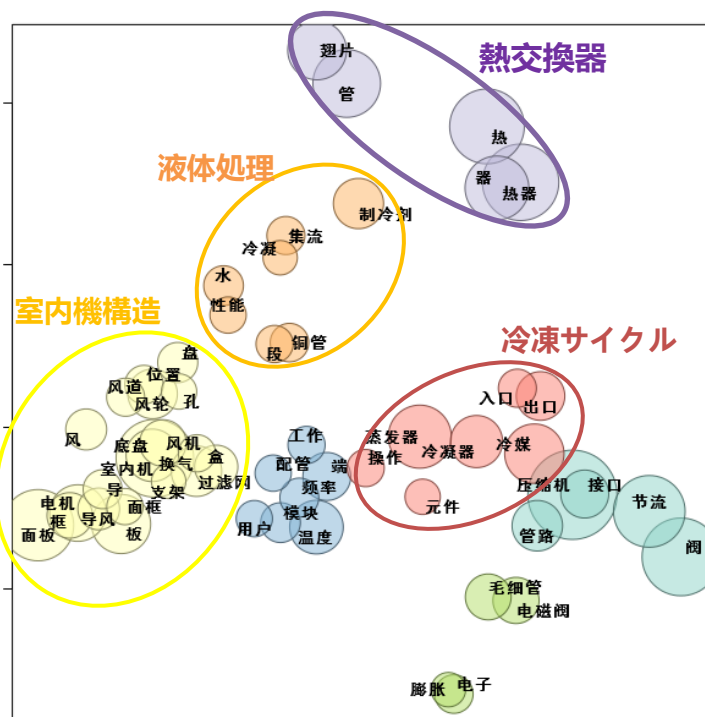
- 統計分析（出願件数の推移確認程度）+ 外部変数を使ったコレスポネンス分析  
⇒ 経年の技術コンセプトの変化確認も可能

## 分かること

- ①2006年～出願が増え始めたが、周辺技術中心
- ②2008～2010年の増加では共通技術の出願が多い
- ③2011年から「冷媒関連」へ重点的に出願されている

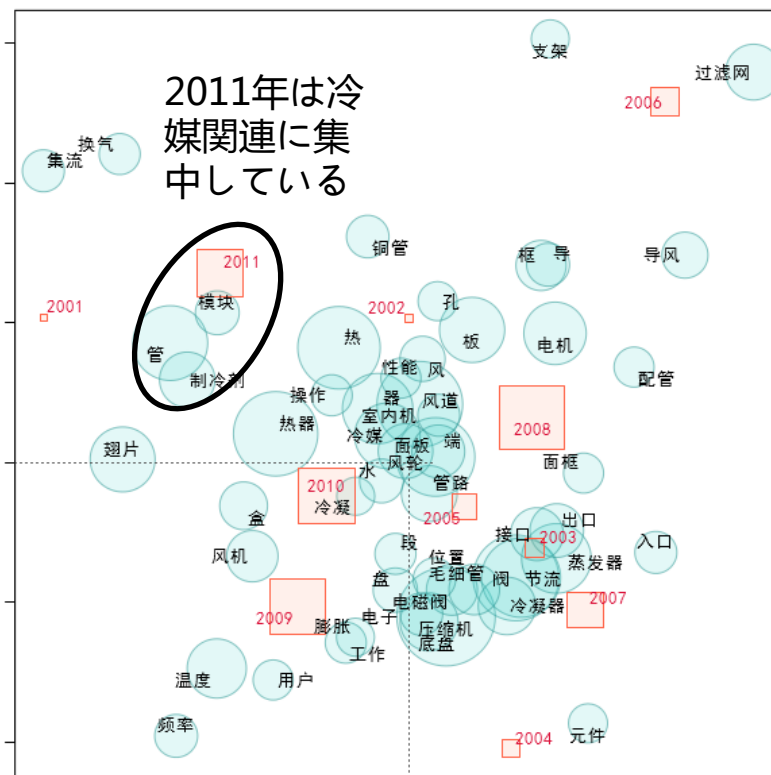
## 空調A社の出願分析





近10年のデータの対応分析：  
出願は2008～2010年で、  
室内機、蒸発器、膨張弁に  
関する共通技術に集中

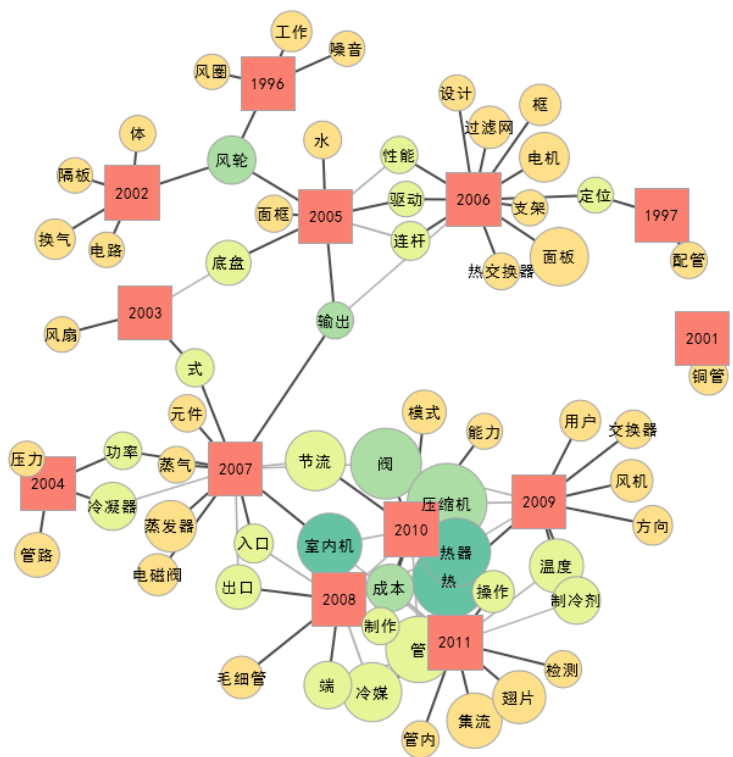
キーワードの類似度を計算した  
多次元尺度の結果、技術分  
類が確認できる



KHCoderの利用により、『年度別の出現頻度、技術コンセプト』を可視化可能、



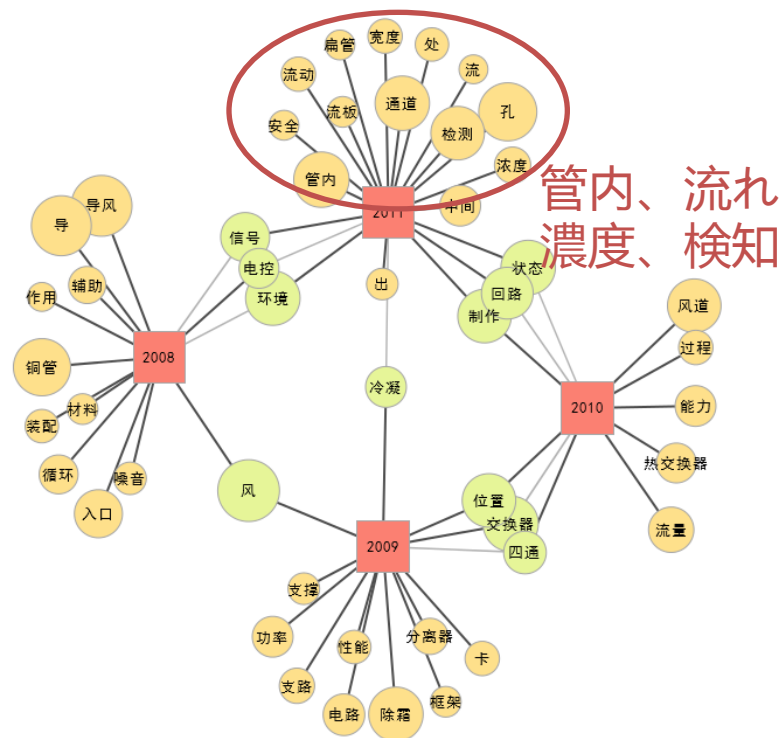
A社萌芽技術を探索できる可能性がある



A社年度別  
出願のコンセプト語抽出



近4年  
抽出



出現頻度「中」  
のコンセプト語

萌芽キーワードを有すると思われる  
2011年の出願に対して再度詳細分析実施

KW

室内機 管内冷媒 濃度 検知  
安全 検知装置 信号 . . .

関連のコンセプト語のかたまりと、それらの  
キーワードのネットワークを確認でき

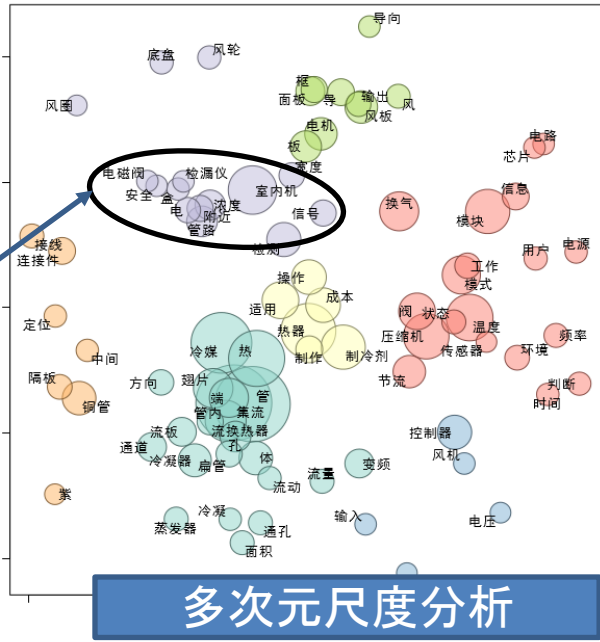
可燃性冷媒の室内機

の開発をしているのでは、との予測ができる

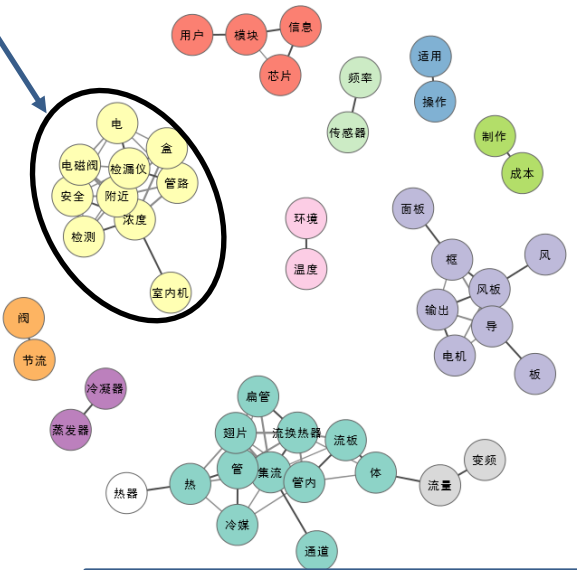
空調冷媒に  
プロパン使用可に



2012年に可燃性冷  
媒使用標準を制定



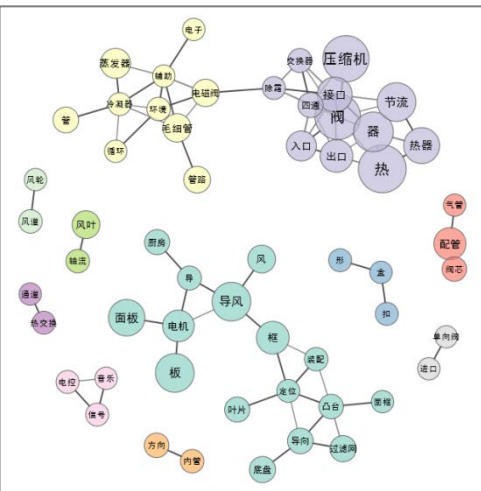
多次元尺度分析



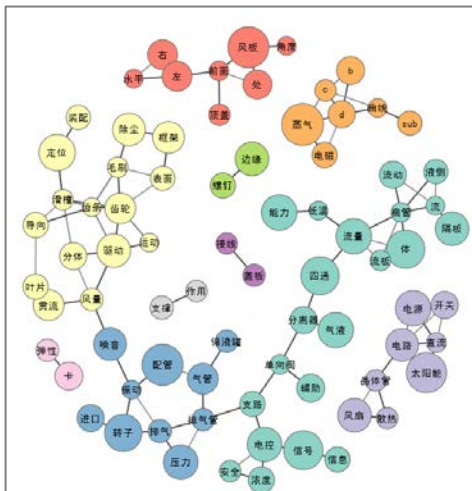
ネットワーク分析

## ③ツールの 組み合わせ分析

## KHCoder



出現頻度「大」



出現頻度「中」

最大/最小出現頻度の取捨選択はできるが、最大/最小/中段の何れしか取れないため、**萌芽KW抽出のための「中段の削除」**  
**+ 最小出現数制限**ができない

※最小出現頻度を極端に小さく設定すると描画時間が極大化

## Rostnat

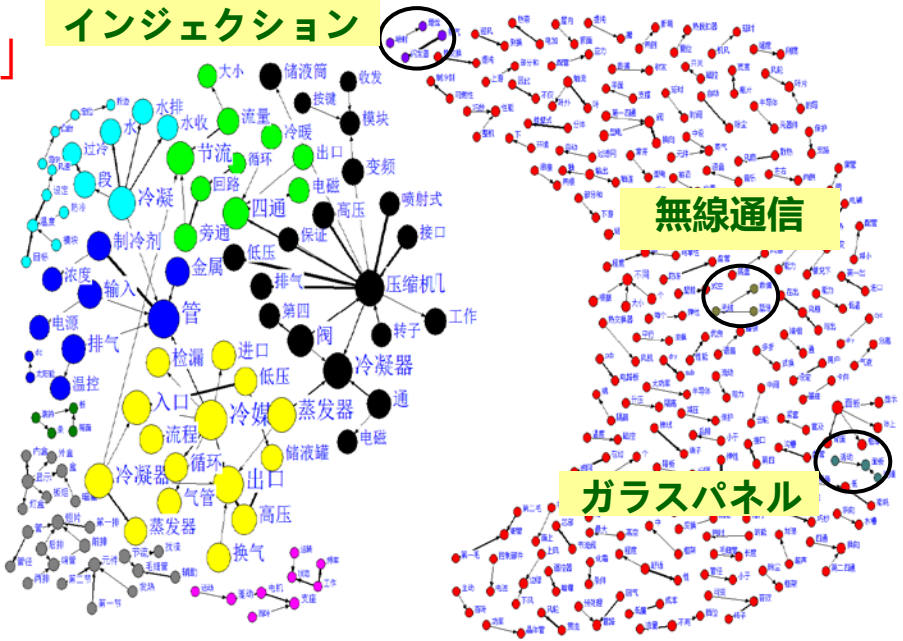
頻度「大」 not 頻度「中」 and 頻度「小」



Rostnat (処理速度最大) で、

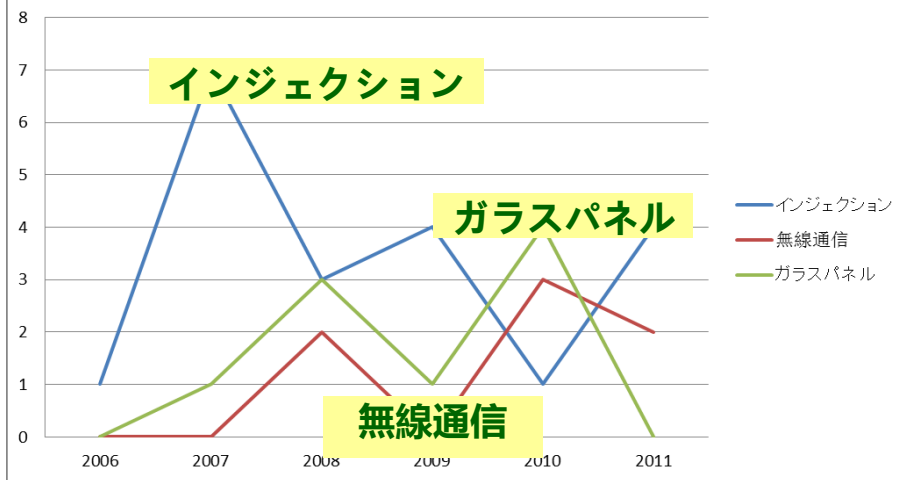
- ・ **最小出現頻度3回以上**
- ・ **出現頻度「中」を**描画しない
- ・ **頻度大の部分**を保留

で、萌芽KWとの関連性の可視化 **迅速**に実現可能





技術キーワードと関連する出願推移



インジェクション、無線通信とガラスパネル、3つの萌芽キーワードを抽出し、それぞれについて母集合の出願傾向を調査

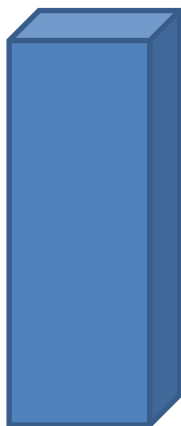
トータル件数で全体に占める割合は少ないが、出願は継続的且つ毎年増えている



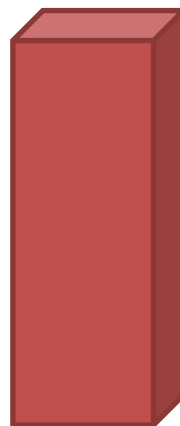
今後関連技術を搭載する商品が出る**可能性がある**と予測

商品化

ガラスパネル  
床置きエアコン  
2012年 公開

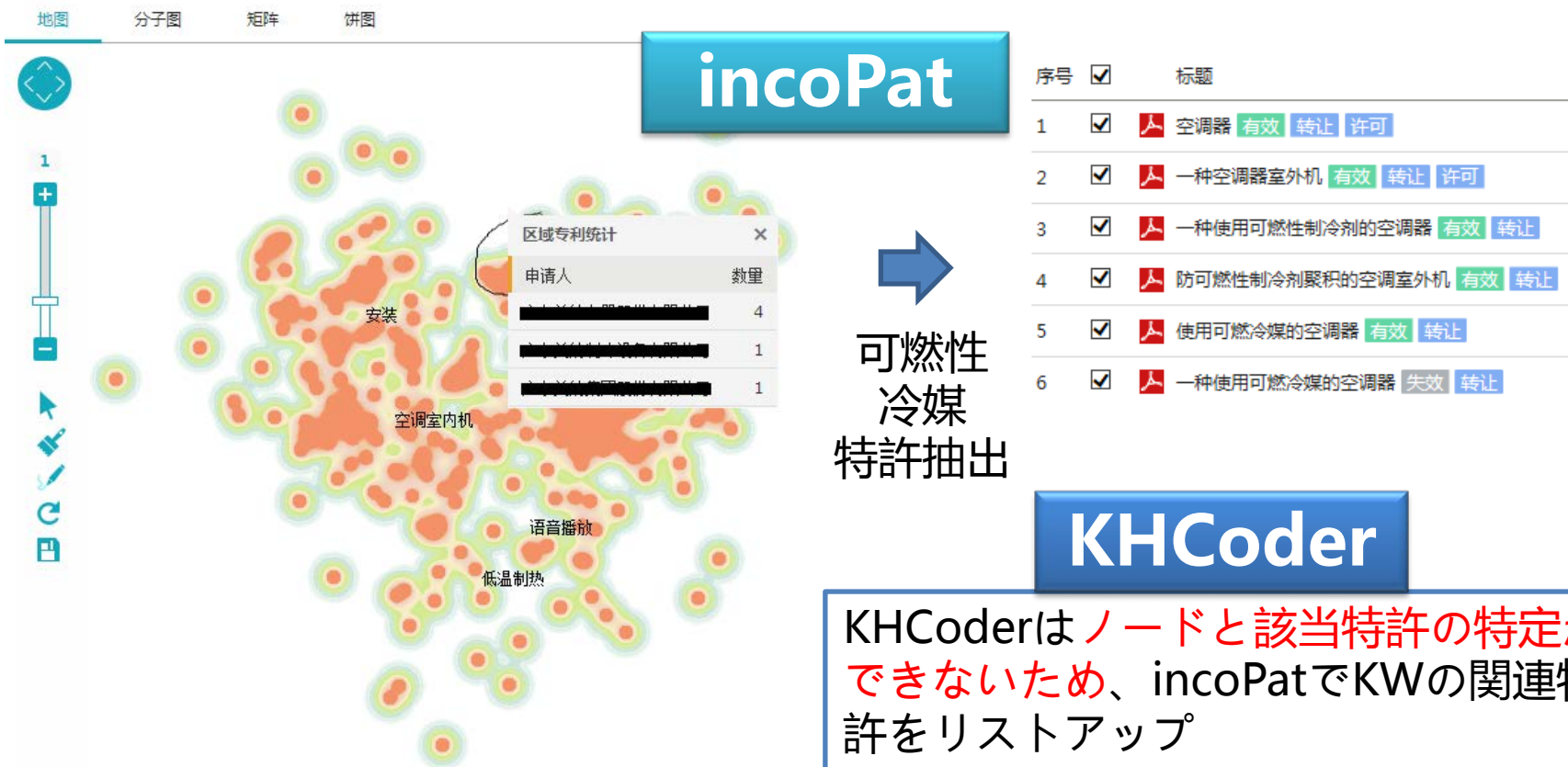


Wifi搭載  
2012年 公開



インジェクション圧縮機  
2013年 公開

incoShare社の商用特許データベースincoPatのテキストマイニング機能を利用し、KHCoderと組み合わせた分析をする



他のツールを組み合わせ、KHCoderを機能補完して分析することで、特許出願から見た技術動向分析を行うことができる可能性は高い

## ~ KHCoderを利用した中国特許分析 ~

- 中国語特許分析の可能性検証した結果、  
分かち書き精度、操作性、可視化の種類において  
現在使われているツールと比べて同等以上の機能がある  
ことを検証できた
- 萌芽キーワードを探索できる
- KHCoderと他のツールの組み合わせにより、より実用的な分析が実現可能

**ご清聴ありがとうございました**

**谢谢大家！**