

中国語テキストマイニングを用いた 中国特許出願動向分析

第12回情報プロフェッショナルシンポジウム
2015年12月11日

松下電器研究開発（蘇州）有限公司

花王株式会社

○袁 方
小林 義典
安藤 俊幸

■背景

■①テキストマイニング分析ツールの選定

- 分詞ツールの検証
 - 分析（可視化）ツール調査
 - 結論
-

■②中国語テキストマイニングでの分析事例

- 条件設定
 - データの整理と分析
 - 空調関連三社の可視化
 - テキストマイニング結果からわかること
-

■③中国語と英語でのテキストマイニング結果の比較

- 中国語と英語での分析対比（事例）
 - 中・英分析の比較結果考察
-

■まとめ

< 現状 >

- ・ 中国語原文のテキストマイニングツールは未評価
- ・ 中国語特許のテキストマイニング分析時、
英文などの他言語で実施するしかない。
(訳文の精度、など不安要素あり)



中国語原文に対して **テキストマイニング + 可視化**
可能なツールが必要



「課題」

- ・ 分詞の精度、及び分詞ツール
(統計機能、フィルター機能、名寄せ機能)
- ・ 中国語対応可視化ツール (文字化け、操作性、利便性)

①テキストマイニング 分析ツールの選定

- ・ 分詞ツール
- ・ 可視化ツール

分詞ツールの評価(1)

分詞 (分かち書き、形態素分析) の比較

(対象 : CN201110176853.7/要約)

本发明公开了一种空调及其空调出风口装置,其包括有:出风口主体及至少一个导风板,所述导风板安装于出风口主体内,其中,所述出风口主体具有第一侧边及第二侧边,所述第一侧边、所述第二侧边分别处于所述导风板的上方、下方,所述第一侧边和/或所述第二侧边为斜面,且所述第一侧边、所述第二侧边的倾斜方向与所述导风板的倾斜方向相同。边设置为与所述导风板的倾斜方向相同的斜面,而形成倾斜的风道,风沿该风道吹至地面,使室内温度分布均匀,具有较高的舒适性;同时,导风板的外侧、内侧均为迎风区,缩小了导风板的外侧与内侧的温度差,避免所述导风板上有凝露现象。

No.	数据名称	所有
1	温度差	1
2	外侧	1
3	凝露现象	1
4	空调	1
5	空调出风口装置	1
6	迎风区	1
7	室内温度分布	1
8	斜面	1
9	出风口主体	1
10	出风口装置	1

PatentMap EXZ

本/rz 发明/n 公开/v 了/ule 一/m 种/q 空调/n 及其/cc 空调/n 出/vf
 风口/n 装置/n /wd 其/rz 包括/v 板/wng /wp 出/vf 风口/n 主体/n
 及/cc 至少/d 一个/mq 板/wng 风/n 板/ng /wd 所/usuo 述/vg 板/wng
 风/n 板/ng 安装/v 于/p 出/vf 风口/n 主体/n 内/f /wd 其中/rz /wd
 所/usuo 述/vg 出/vf 风口/n 主体/n 板/wng 第一/m 侧/q 边/n 及/cc
 第二/m 侧/q 边/k /wd 所/usuo 述/vg 第一/m 侧/q 边/n /wd
 所/usuo 述/vg 第二/m 侧/q 边/k 分别/d 位于/v 所/usuo 述/vg 板/wng
 风/n 板/ng 的/ude1 上方/f /wd 下方/f /wd 所/usuo 述/vg 第一/m
 侧/q 边/k 和/p /w 或/c 所/usuo 述/vg 第二/m 侧/ng 边/d 为/v
 斜面/n /wd 且/c 所/usuo 述/vg 第一/m 侧/q 边/n /wd 所/usuo
 述/vg 第二/m 侧/q 边/k 的/ude1 相同/vn 方向/n 与/p 所/usuo 述/vg
 板/wng 风/n 板/ng 的/ude1 相同/vn 方向/n 相同/v /w 本/rz 发明/n
 通过/p 将/d 所/usuo 述/vg 第一/m 侧/q 边/k 和/p /w 或/c 第二/m
 侧/ng 边/d 设置/v 为/v 与/p 所/usuo 述/vg 板/wng 风/n 板/ng 的/ude1

NLPIR

集計結果

侧边	10
导风板	8
导风	8
风板	8
一侧	5
风口	4
倾斜	4
边	3
主体	3
出	3
方向	3
斜面	2
外侧	2

PatAnalyzer

侧边	10
导风板	7
第二侧边	5
倾斜	4
出风口	3
主体	3
方向	3
相同	2
风道	2
温度	2

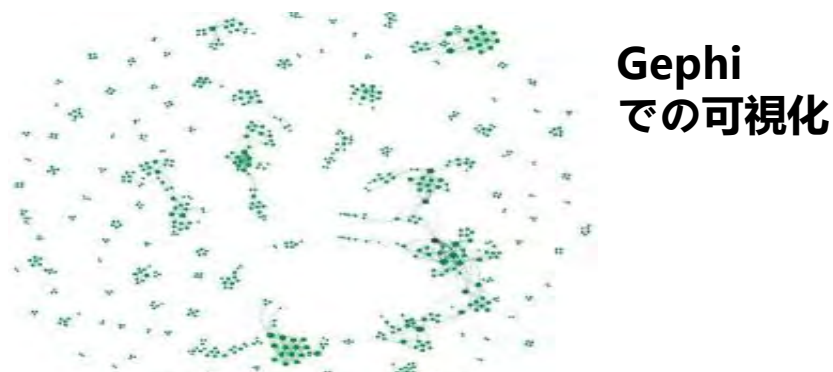
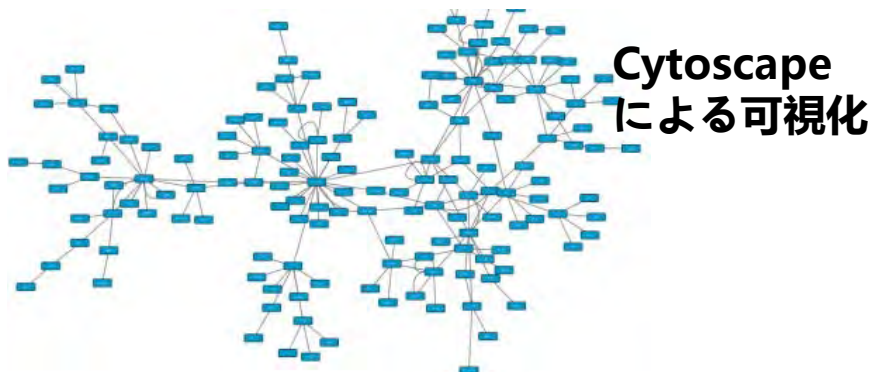
Rostnat

分詞ツールの評価(2)

ソフト名	出所	TM	頻度統計 ランキング	フィルター (名寄せ)	備考
MTMineR(MLTP)	同志社大学 金明哲教授	○	△ 難	△ 難	Rと連動で分析も可能 専門家向き
NLPIR(ICTCLAS)	中国科学研究院	○ 精度高	×	×	形態素子の認識度 最も高い
RostCM(Rostnat)	武漢大学インターネット 科学研究C	○	○	○ 名寄せなし	ビッグデータ対応可能、 分詞速度速い
PatentMap EXZ	インパテック社	△ 精度低	○	△ フィルターなし	分詞精度△ (統計向きではない)
IKAnalyzerNet	林 良益	○	×	×	データ少量でフリーズ (要約約xx件<)
PatAnalyzer (IKAnalyzerNet)	花王 安藤 俊幸	○	○	○	IKAnalyzerのエンジンを使 った自作ソフト

**統計用分詞精度、大容量データ対応 (1万件以上)、名寄せ
⇒ 「Rostnat」、 「PatAnalyzer」 が最適**

ソフト名	出所	データ分析	可視化	備考
Cytoscape	U.S. National Institute of General Medical Sciences	×	○	共起ネットワークの可視化可能
Gephi	Gephi team	×	○	共起ネットワークの機能は強い が、動作環境の要求も高い
RostCM (Rostnat)	武漢大学インターネット 科学研究センター	○ ビッグデータ 対応	○	可視化はNetdrawで実現可
Ucinet6 (Netdraw)	Analytic Technologies	○	○	Excelで事前処理が必要



**データ可視化機能（共起ネットワーク分析）
⇒ 「Rostnat」 「Ucinet6」 が好適**

テキストマイニング分析（分詞～統計～可視化）ツール：

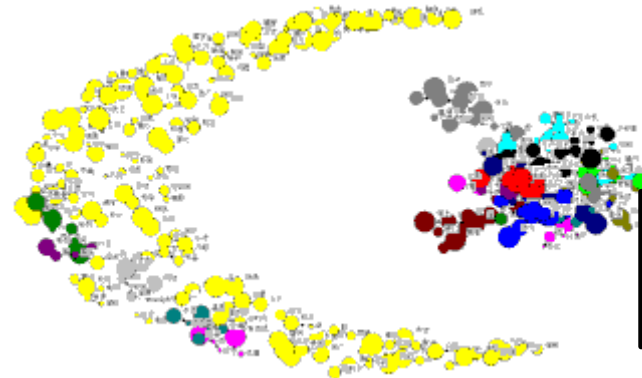
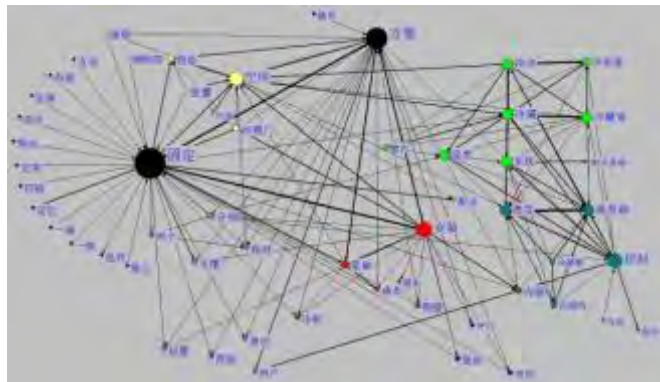
① 「Rostnat」

② 「PatAnalyzer + Netdrawの組み合わせ」

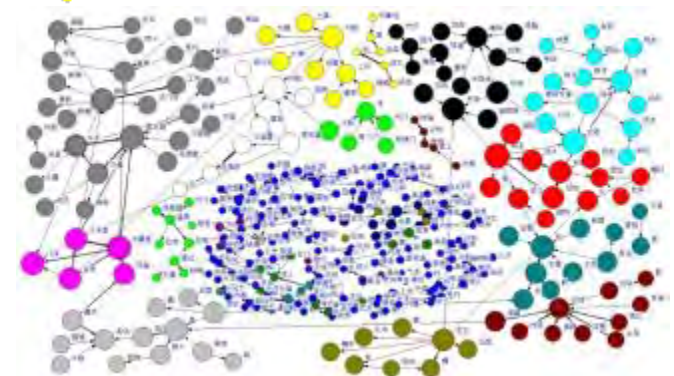
がベストと評価



RostNatによる
可視化結果



PatAnalyzer +
Netdrawによる
可視化結果



②中国語テキストマイニングでの分析事例

- ・空調分野の分析事例

事例) 空調分野3万件の分析(1)

条件

中国語公報DB : PatentSQUARE

検討データ : 公開日1985年~2015年6月30日までの全件

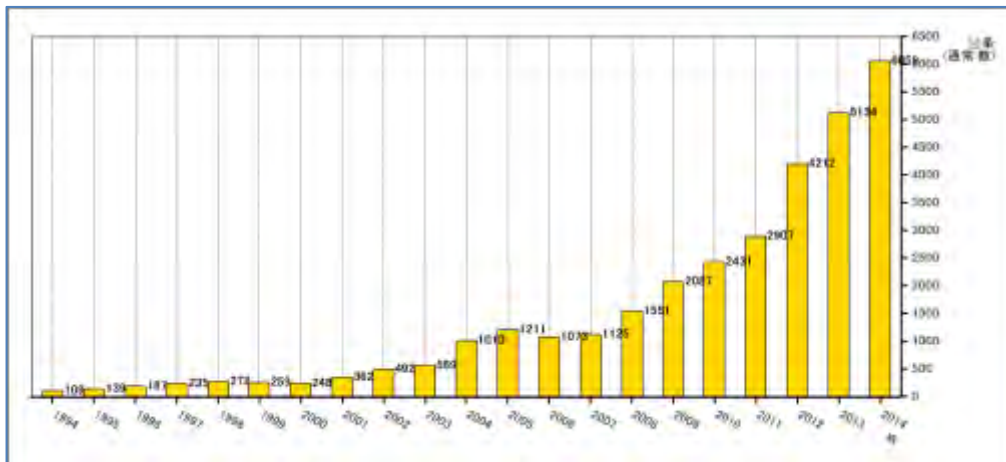
TM内容 : **要約**

IPC : F24F 01/00+F24F 11/00+F24F 13/08+F24F 13/10+F24F 13/14+F24F 13/15+ F24F 13/16+F24F 13/18+F24F 13/20+F24F 13/22+F24F 13/24+F24F 13/26+ F24F 13/28+F24F 13/30+F24F 13/32

抽出件数 : 32,333件

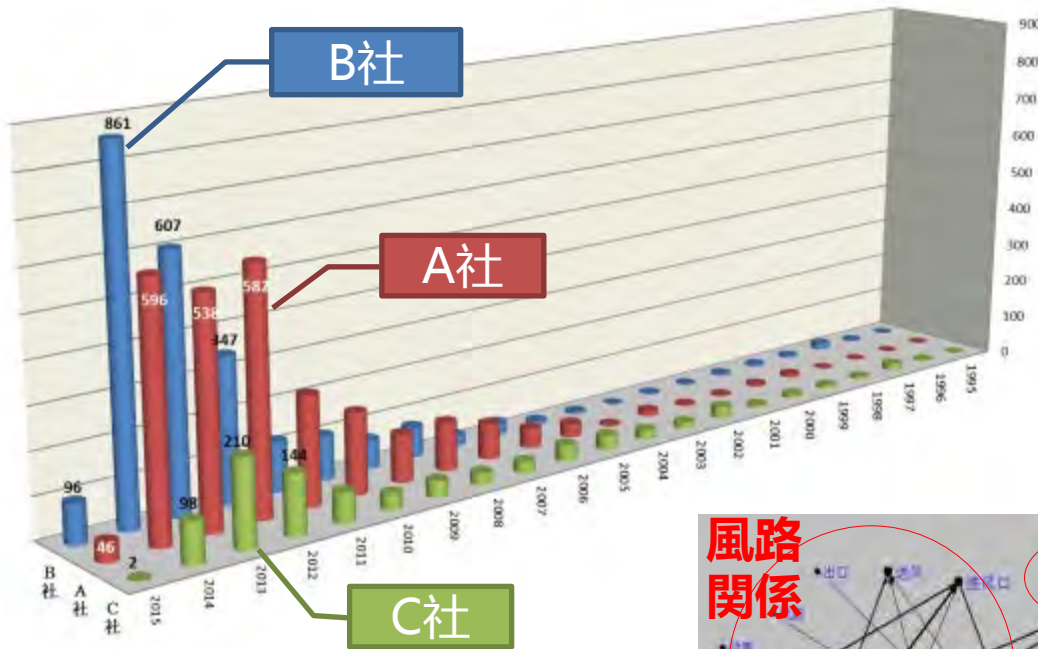
式No.	登録件数	検索項目	条件式
S001	34,134	IPC	F24F 01/00+F24F 11/00+F24F 13/08+F24F 13/10+F24F 13/14+F24F 13/15+ F24F 13/16+F24F 13/18+F24F 13/20+...
S002	6,110,428	公告・登録日	:20150630
S003	4,730,809	公開日	:20150630
S004	9,155,402	論理式	S002+S003
S005	32,333	論理式	S001*S004

(PRE処理) 出願人名寄せ : TECRES (中国版PatentMap EXZ)
件数 > 2万件 でも処理速度問題なし(Excel×)



中国空調分野の出願件数推移

事例) 空調分野3万件の分析(2)



TM処理速度

処理速度：「Rostnat」

要約件数：32,333件

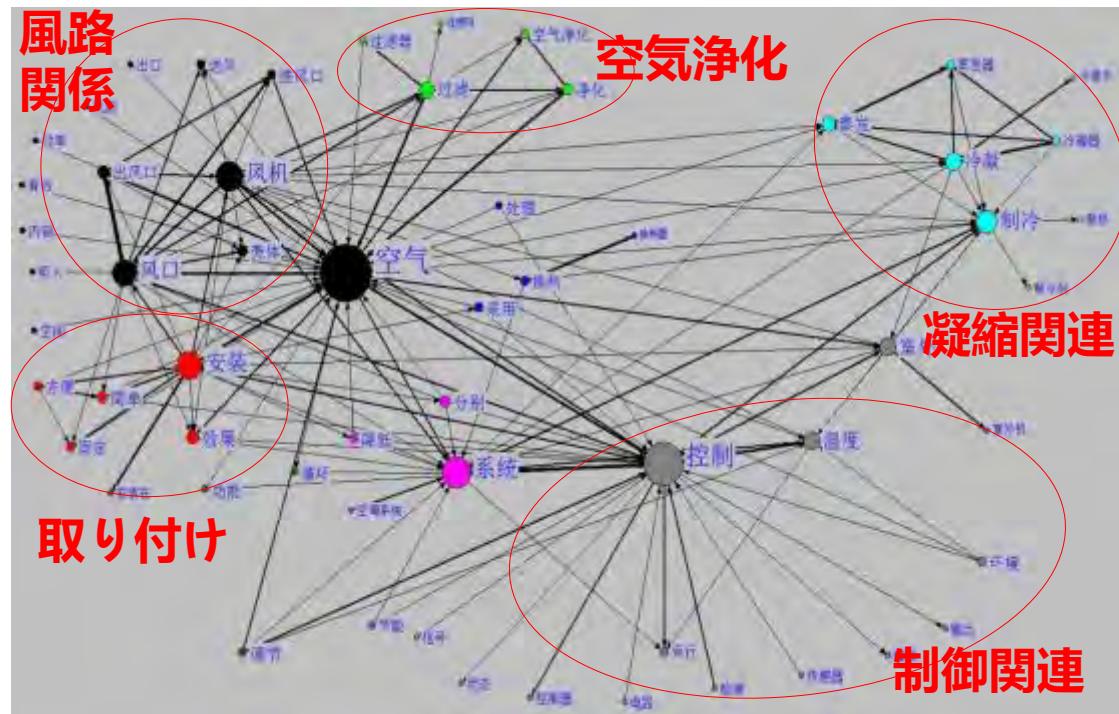
漢字：7,445,774文字

分詞&統計処理：170秒

フィルタリング：3秒

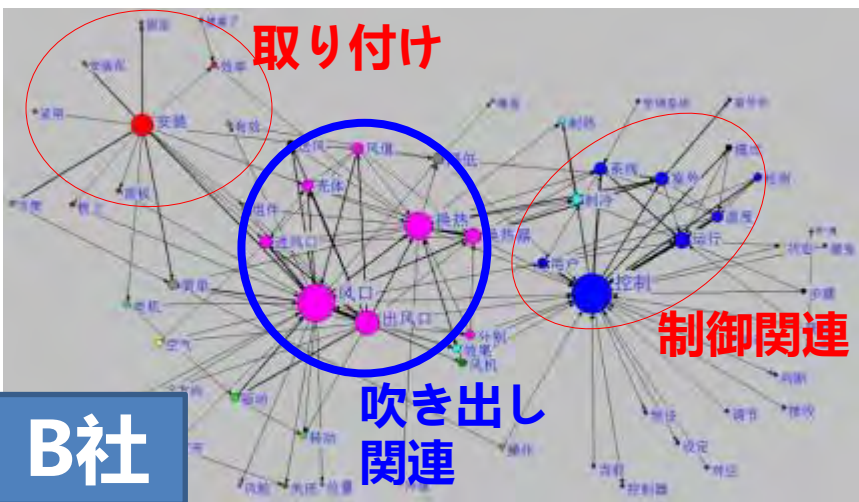
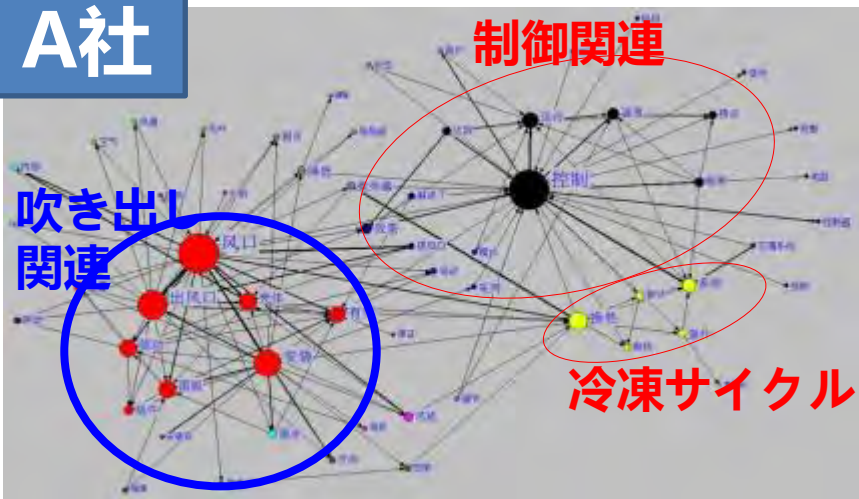
可視化：32秒

中国の空調シェアTOP3
の出願動向
(合計販売シェア6割超)

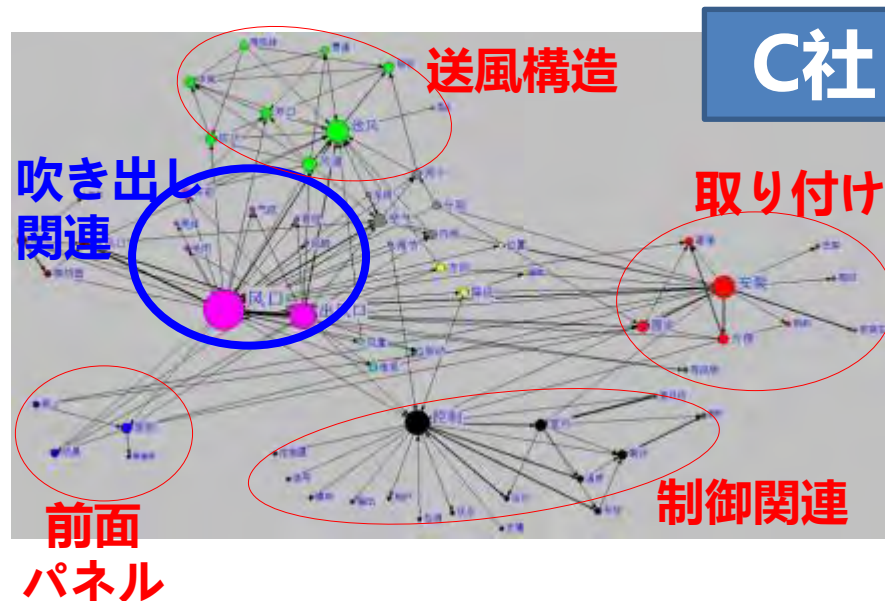


中国主要メーカー3社のPFの特徴抽出

A社



B社



C社

三社とも『吹き出し関連』の特徴語多数

中国企業は『吹き出し口の配置構成』、
『羽根の動き』に注力していると推定

**③中国語と英語での
テキストマイニング
結果の比較評価**

背景

- ・日本人からすれば「英語TM」の方が利用しやすいはず
- ・解析ツールも充実

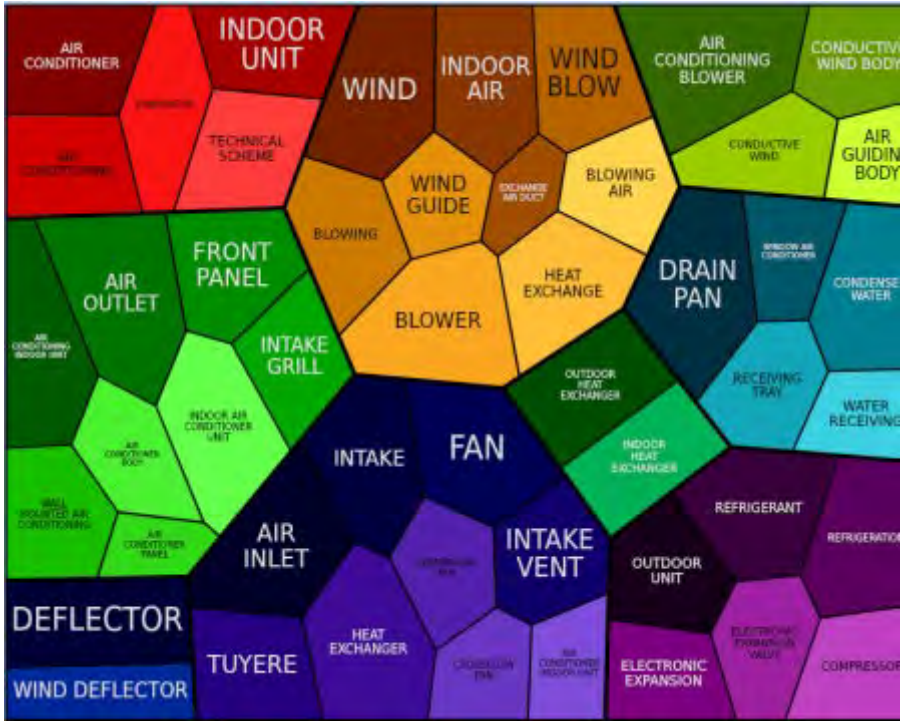


「英語TMと中国語TMで
分析結果に差異がみられるか？」

- 分析結果の対比を実施
 - ・可視化結果
 - ・頻出形態素

英語TM分析の結果（空調/C社）

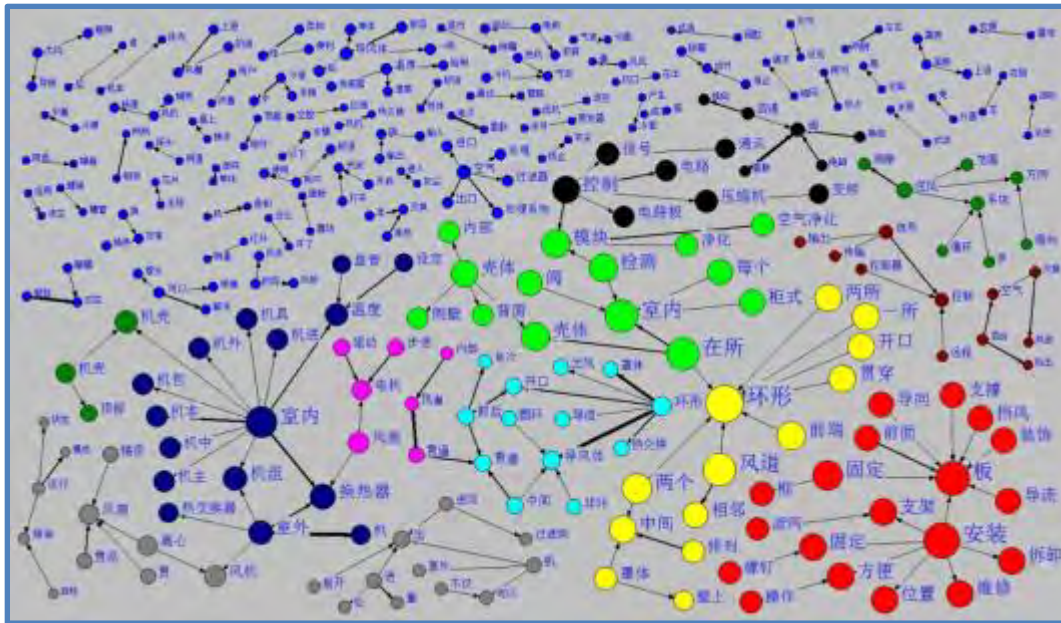
DB : Orbit.com
件数 : 869件
対象 : 要約



コンセプト	Value	コンセプト	Value	コンセプト	Value
AIR CONDITIONER	614	AIR CONDITIONER INDOOR UNIT	62	CONDUCTIVE WIND	28
INDOOR UNIT	341	DRAIN PAN	60	OUTDOOR MACHINE	28
AIR CONDITIONING	270	COOLING MODE	60	WINDOW AIR CONDITIONER	27
TECHNICAL SCHEME	269	BLOWING AIR	58	MOTIVATED WIND	26
AIR OUTLET	211	AIR INTAKE	57	AIR CONDITIONER PANEL	26
WIND	206	WIND DEFLECTOR	56	BLOWING PERFORMANCE	26
AIR INLET	184	AIR CONDITIONING BLOWER	50	ANNULAR HEAT	26
FAN	178	WIND GUIDE	50	AIR CONDITIONING AIR OUTLET	26
INTAKE	160	ELECTRONIC EXPANSION VALVE	50	WALL MOUNTED AIR CONDITIONING	25
OUTDOOR UNIT	153	WIND DIRECTION	49	COMPRESSOR FREQUENCY	25
COMPRESSOR	150	WIND OUTLET	48	AIR CONDITIONER TECHNICAL FIELD	25
EVAPORATOR	141	AIR BLOWING	45	AIR CONDITIONING OPERATION	25
HEAT EXCHANGER	134	ELECTRONIC EXPANSION	44	AIR GUIDING BODY	24
AIR CONDITIONING INDOOR UNIT	118	HOUSEHOLD APPLIANCE	44	VERTICAL AIR CONDITIONER	24
BLOWER	118	CROSSFLOW FAN	43	INDOOR AIR CONDITIONER UNIT	24
INTAKE VENT	116	OUTDOOR HEAT EXCHANGER	42	AIR CONDITIONING BLOWOUT	23
TECHNICAL PROPOSAL	113	REAR OPENING	42	ROOM AIR UNIFORMITY	22
REFRIGERANT	107	INDOOR WIND	40	UPRIGHT AIR CONDITIONER	22
CONDENSATION	103	USER COMFORT	40	AIR CONDITIONING PANEL	22
DEFLECTOR	99	EXCHANGE AIR DUCT	39	INDOOR UNIT EVAPORATOR	22
DUCT	98	WIND BODY	38	AIR CONDITIONER AIR CONDITIONER	21
REFRIGERATION	94	INDOOR HEAT EXCHANGER	38	AIR CONDITIONER BLOWER	21
INDOOR AIR	90	HYBRID WIND	35	MIXED AIR OUTLET	20
BLOWING	89	EXCHANGE WIND	34	INDOOR AIR UNIFORMITY	19
EQUIVALENT SUBSTITUTION	87	AIR CONDITIONER BODY	34	WALL MOUNTED AIR CONDITIONER	19
FRONT PANEL	82	AIR EXCHANGE	34	VERTICAL AIR CONDITIONING	18
HEAT EXCHANGE	78	GUIDE BODY	34	BLOW VELOCITY	18
TUYERE	74	VOLUTE	34	INTAKE SHAPED WIND	17
WIND BLOW	70	INTAKE GRILL	32	INTERMEDIATE WIND	17
CENTRIFUGAL FAN	68	OUTDOOR FAN	31	BREEZE BLOWOUT	16
AIR DUCT	68	WIND INLET	30	AIR CONDITIONER QUANTITY	16
CONDENSED WATER	66	CONDUCTIVE WIND BODY	28	REGULAR POLYGONAL RING	16
COMFORT	66	HEAT EXCHANGE AIR	28	PRODUCTION EFFICIENCY APPLICABILITY	13

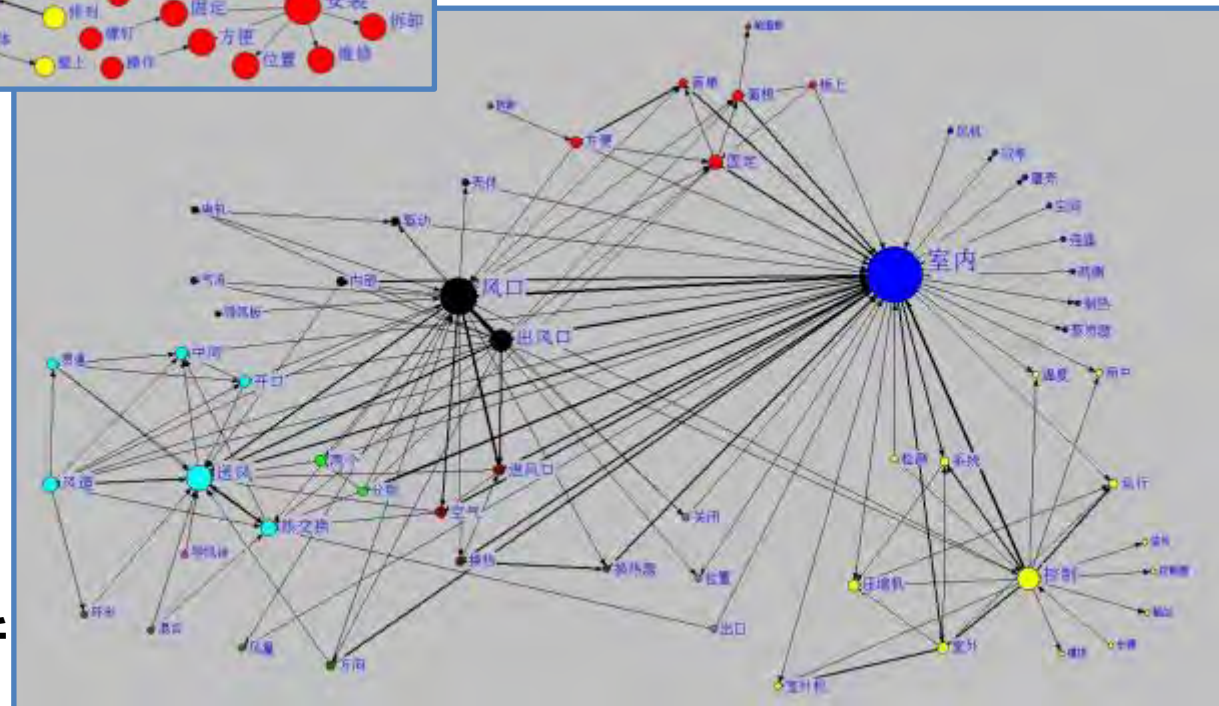
Orbitの分析





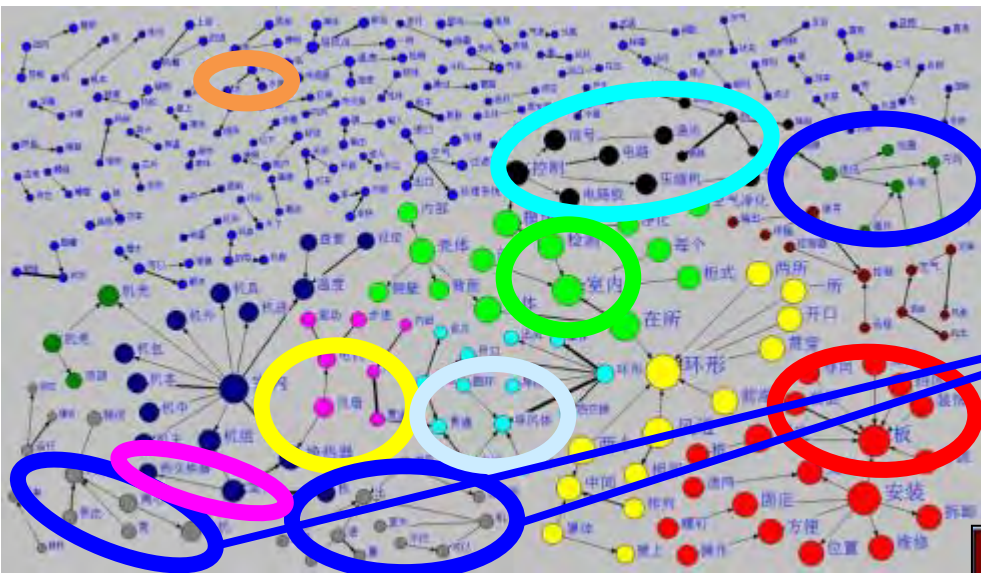
DB : PatentSQUARE
件数 : 869件
対象 : 要約

PatAnalyzer +
Netdrawの分析



Rostnatの分析

C社 869件



共通項目

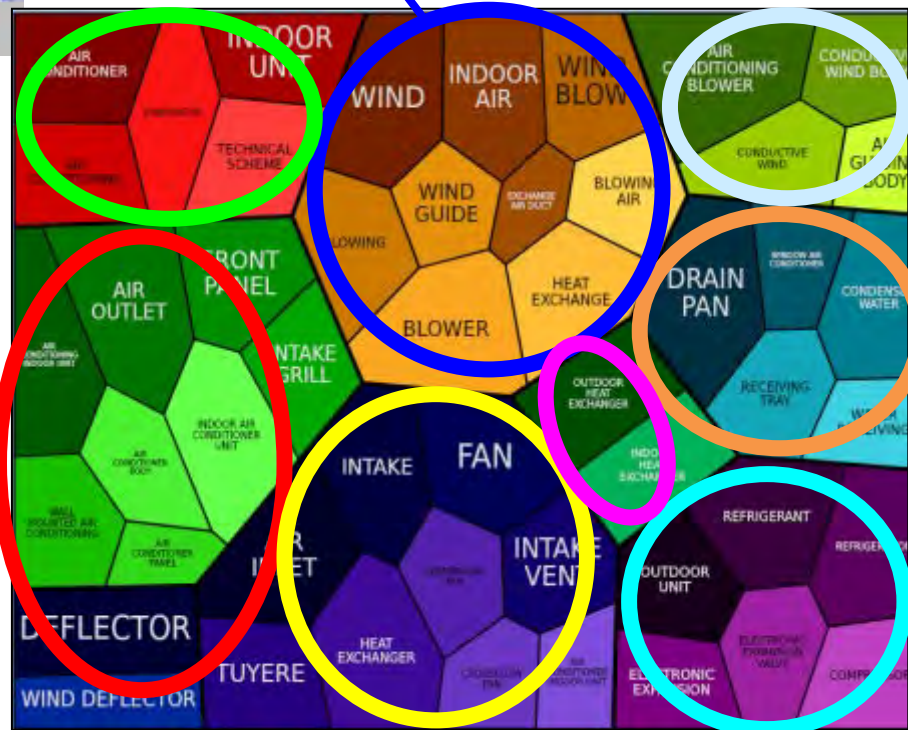
2層目により多くの情報がありそう



相違点：

(中国語分析○、英語分析×)

- 1. 環状配置風路&熱交換器 (特徴技術)
- 2. 空気浄化関連
- 3. 床置きエアコン
- 4. 遠隔操作制御
- 5. 全熱交換器 (換気) 関連



中国語

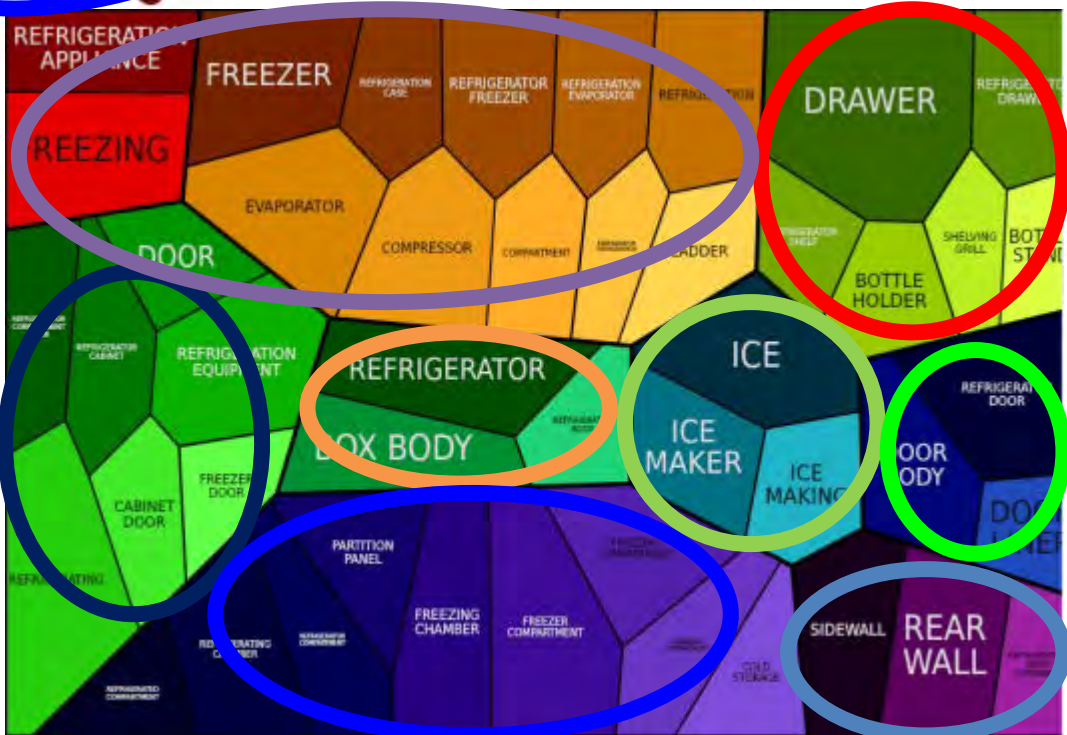
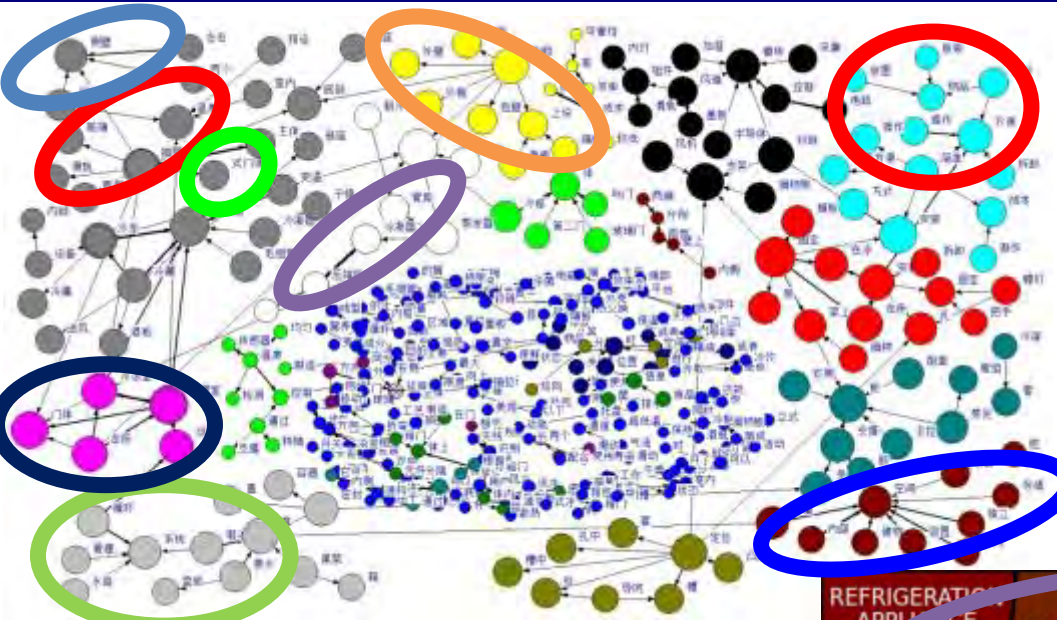
No.	単語	日本語訳	頻度
1	固定	固定	1159
2	抽屉	引き出し	884
3	蒸发器	蒸発器	785
4	内胆	インナー	772
5	安装	取り付け	662
6	空间	空間	554
7	控制	コントロール	553
8	冷柜	冷凍庫	548
9	冰箱门	冷蔵庫のドア	516
10	冷藏室	冷蔵室	493

英語

No.	コンセプト	日本語訳	Value
1	DOOR	ドア	597
2	REFRIGERATOR DOOR	冷蔵庫のドア	588
3	REFRIGERATION	冷凍	533
4	DOOR BODY	ドア本体	511
5	REFRIGERATOR CABINET	冷蔵庫キャビネット	409
6	FREEZER	冷凍庫	409
7	REFRIGERATED COMPARTMENT	冷凍庫	402
8	SIDEWALL	側壁	377
9	REFRIGERATING CHAMBER	冷蔵室	370
10	REFRIGERATOR COMPARTMENT	冷蔵庫区画	338
11	FREEZING CHAMBER	冷凍室	309
12	EVAPORATOR	蒸発器	292
13	FREEZER COMPARTMENT	冷凍庫区画	254
14	RACK	ラック	243
15	DRAWER	引き出し	242
16	COMPRESSOR	コンプレッサー	242
17	FOOD	フード	227
18	BOX BODY	ボックス本体	219
19	CONDENSATION	凝縮	213
20	COMPARTMENT	区画	211

中国語に比べ英語では翻訳時に複数の同義語、類似表現が使われており『同義語辞書』があれば中国語分析と同等のワード集約ができ有効

C社 1828件



相違点：
 (中国語分析○、英語分析×)
 1. 位置決め
 2. モジュール制御
 3. 固定関係
 4. ガラスドア&センサー
 5. UD (ユニバーサルデザイン) 関連

中国語原文分析

体言（名詞）を修飾するもの

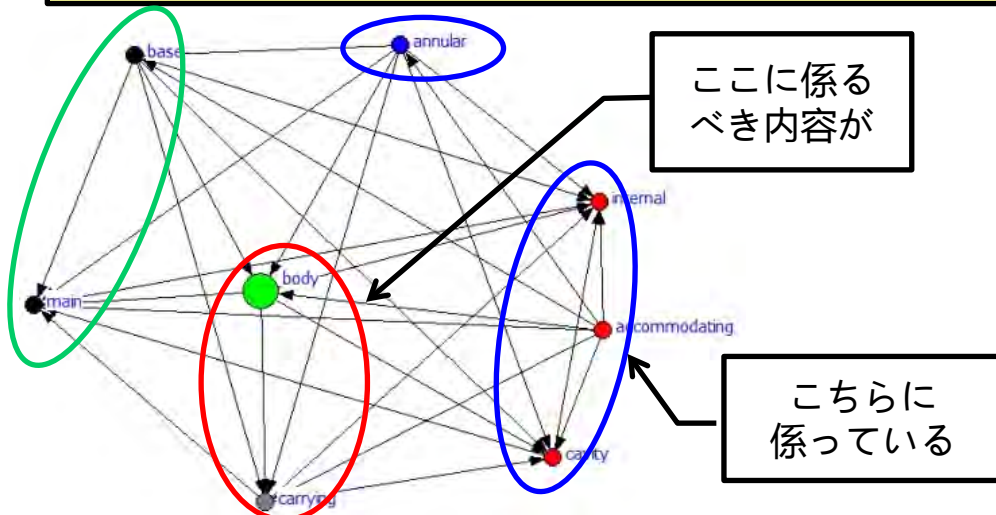
- 『主語と定语（連体修飾語※）の係り受け』が重要

中国語：修飾関係が比較的シンプル 英語：前後からの修飾、複雑

所述基座主体上设置内部具有容置空腔的环形承载体。

前記本体上に、收容穴を備える、環状の搭載物を設置されている。

wherein the base main body is provided with an annular carrying body with an internal accommodating cavity.



『中国語→英語』

翻訳によっては、名詞への修飾関係が複雑化/分散化

分散された修飾語がそれぞれ独立して抽出され、相関関係の認識が難しくなるのではないかと推定

- 1 . **中国語原文テキストマイニング～可視化ツールを抽出**
 - ・ 「Rostnat」 or 「PatAnalyzer+UCINET6」 の組合せ
 - ・ ポートフォリオの概要は認識できることを確認

- 2 . **中国語分析の方が英文分析に比べ概要把握が容易**
 - ・ 英語分析では単語の集約のPREが望ましい
 - ・ 中国語分析では主語 - 定語の認識が容易
(英語分析では修飾関係のあるKWが分散され
全体の認識に対して影響)

ご清聴ありがとうございました

谢谢大家！