

A22 中国語キーワードを用いた特許情報解析

調査精度向上への応用

○花王株式会社	安藤俊幸
インパテック株式会社	有賀康裕
株式会社 I H I	金澤祐孝
株式会社 I H I	乾 智彦

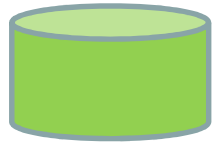
発表内容

1. はじめに
2. 調査対象(風力発電の技術動向調査)について
3. パテントマップEXZによる中国語キーワード抽出
4. 中国語キーワードを用いた特許情報解析
 - ①パテントマップEXZの例
風力発電の技術動向調査
 - ②パテントマップEXZ以外での例
調査用キーワード辞書作成
 - ③英語キーワード解析結果との比較
5. 解析結果の中国特許調査への応用
6. まとめ

1. はじめに

— 中国語キーワード利用の意義 —

中国特許検索

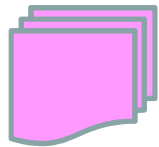


キーワー検索方法

- ・英語
- ・中国語(原語)
- ・日本語(機械翻訳)

} ハイブリッド検索

- ・英語系DB: 機械翻訳が大部分
- ・中国語は異表記も多く、かつ外来語的なものに弱い
- ・従来: 英語KW⇒中国語KW
- ・今回: 中国語KW⇒英語KW
(中国語KW選択が重要)



目的分野の特許公報群から
調査(検索)用キーワード抽出



分野別の調査用
独自辞書作成

目的

① パテントマップEXZでの利用

- ・キーワード等取り込み用ユーザー辞書
- ・パテントマップ**分類用検索式**

② パテントマップEXZ以外での応用

- ・**調査用キーワード辞書**作成(中国語、日本語、英語等)
- ・データベース検索用の中国語検索式作成
(検索精度に影響する**適合率**と、網羅性に関する**再現率の向上**)
- ・各種統計解析(対応分析、多次元尺度法、クラスター分析⁷⁾等)用中国語辞書⁸⁾
- ・テキストマイニングと可視化

使用ツール

使用ツール

○中国語テキスト用データベース

- Questel 社 Orbit.com
- 発明通信社 HYPAT-i

○中国語キーワード抽出、解析

- DataAcross(データ変換ツール)
- パテントマップEXZ

○中国語キーワード関連ツール

- WIPO PATENTSCOPEの多言語検索CLIR(Cross-Lingual Information Retrieval)
- Google翻訳
- 鄭州大学在線英漢-漢英科技大詞典
- 北辞郎(中日辞書)

○英語キーワード解析

- Questel 社 Orbit.com Analysis module

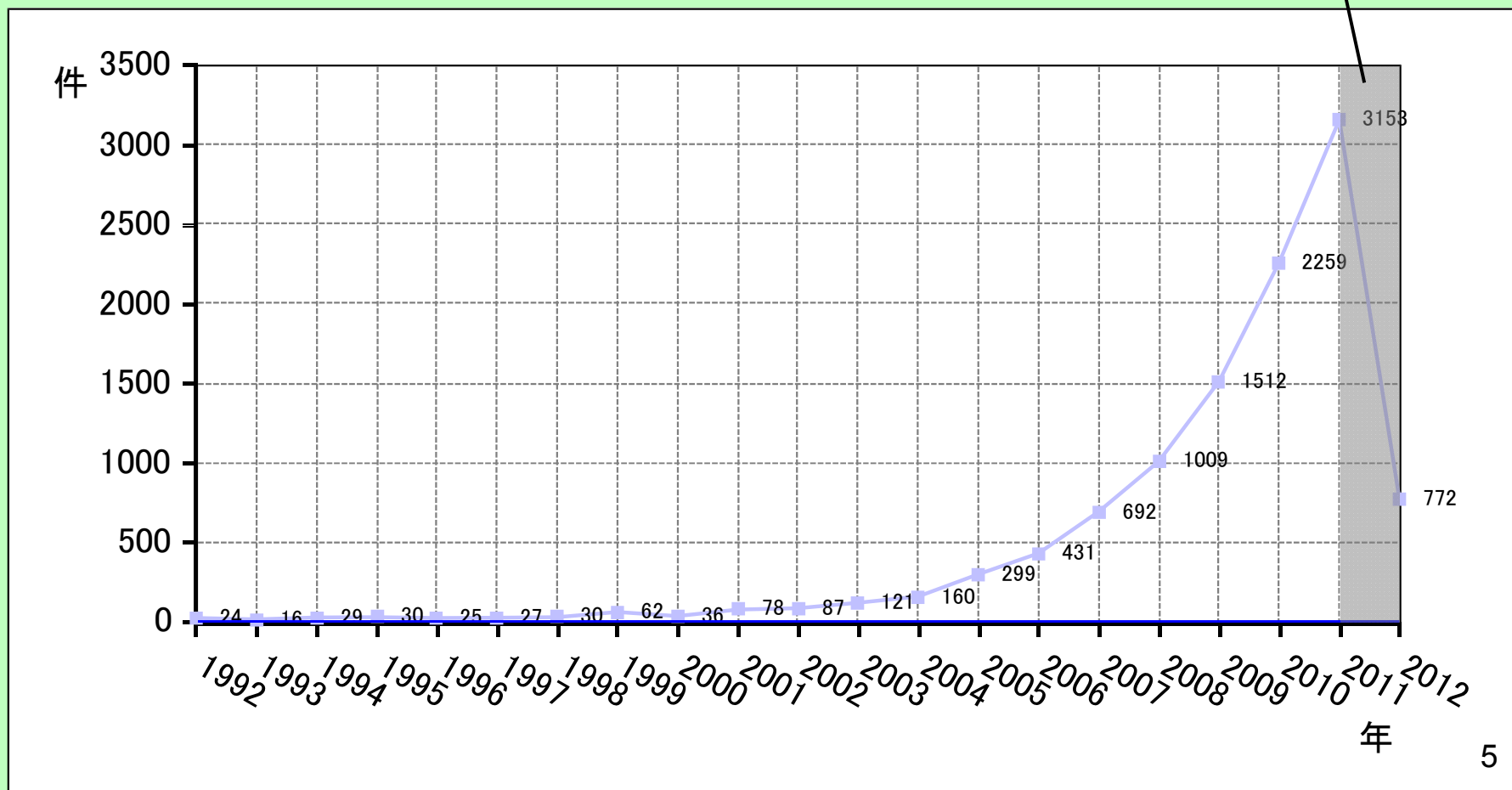
2. 調査対象：風力発電の技術動向調査

Orbit.com (FMAPAT) 検索式

(F03D)/IC AND (CN)/PN AND PD <= 2012-03-31 **11,015件**

風力発電全体の
件数推移時系列マップ

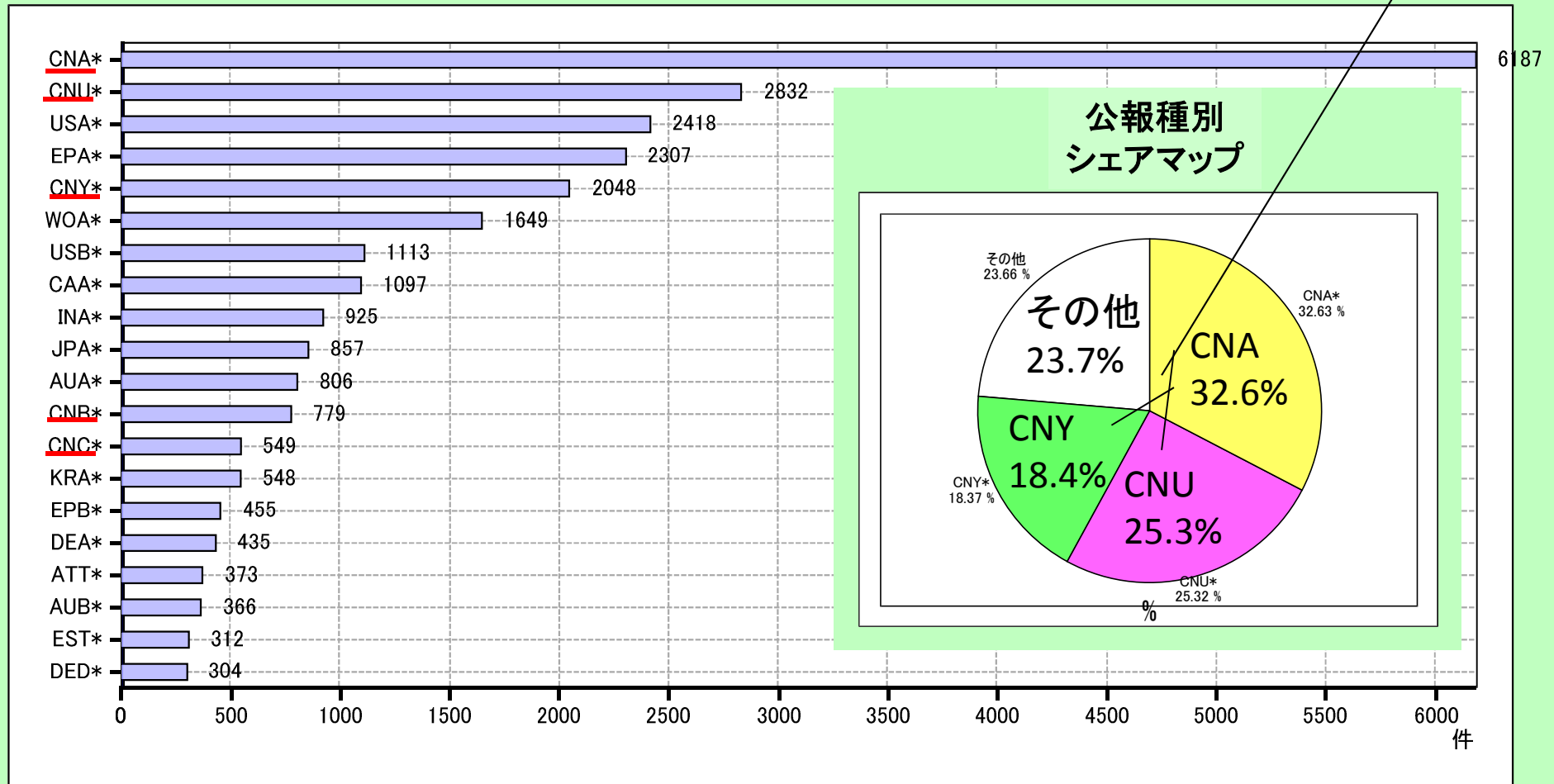
不確定範囲



公報種別ランキングマップ

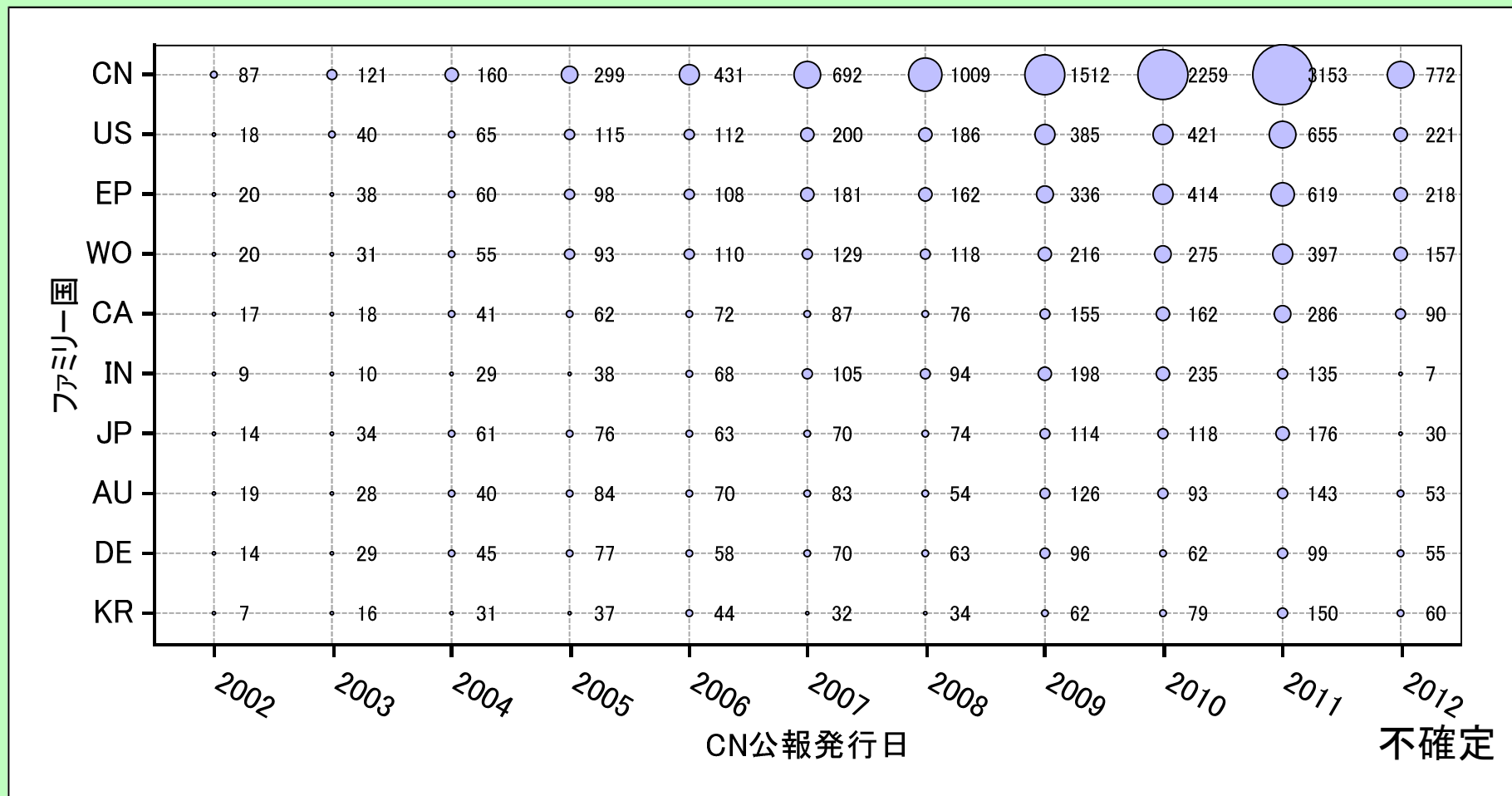
公報種別
件数ランキングマップ

中国のみの**単独出願**: 74.3%
(OrbitのファミリーXPNを別途集計)



公報発行国(ファミリー国)時系列マップ

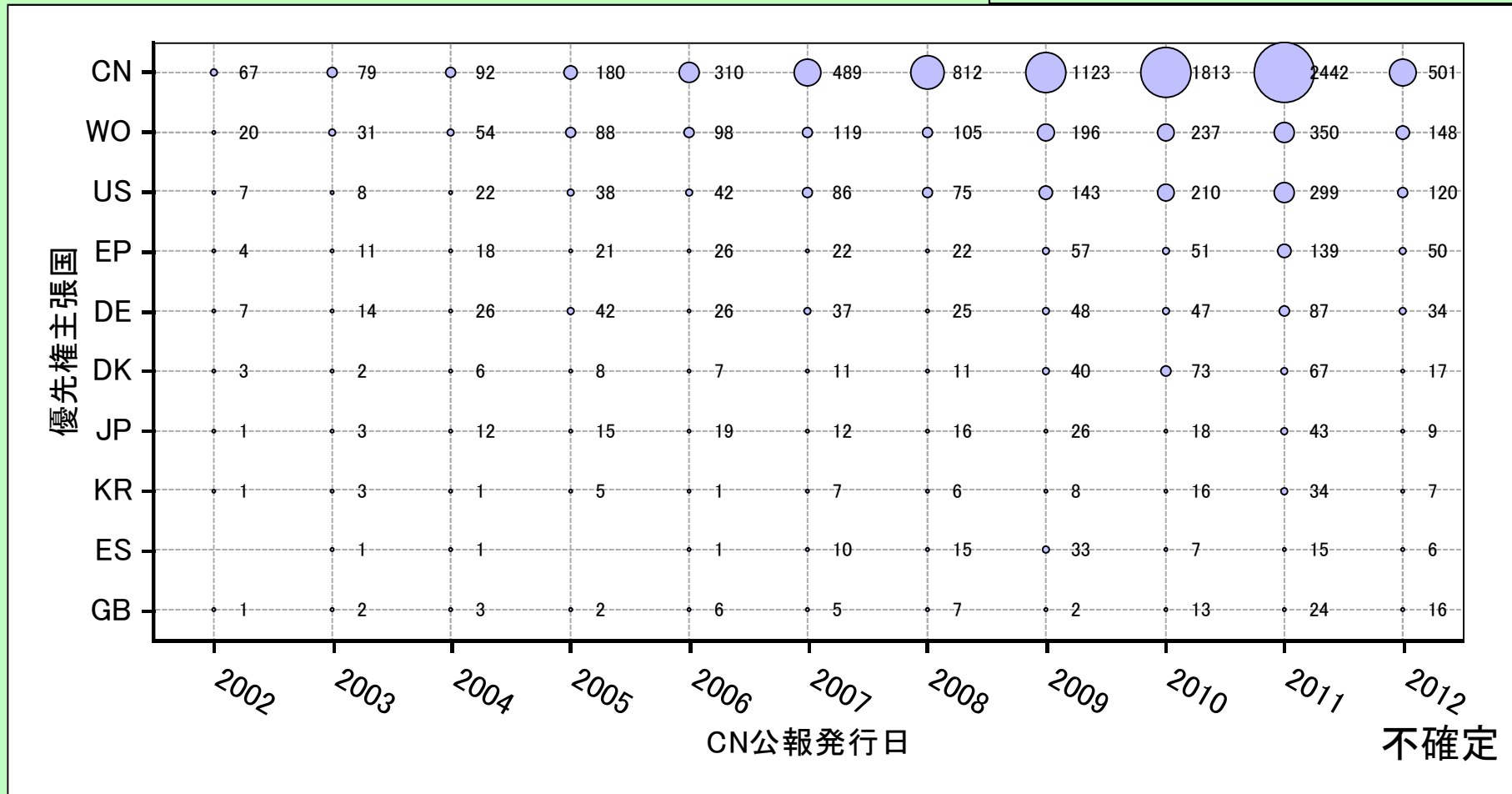
ファミリー国 * CN公報発行日(年)
件数マトリクスマップ



優先権主張国時系列マップ

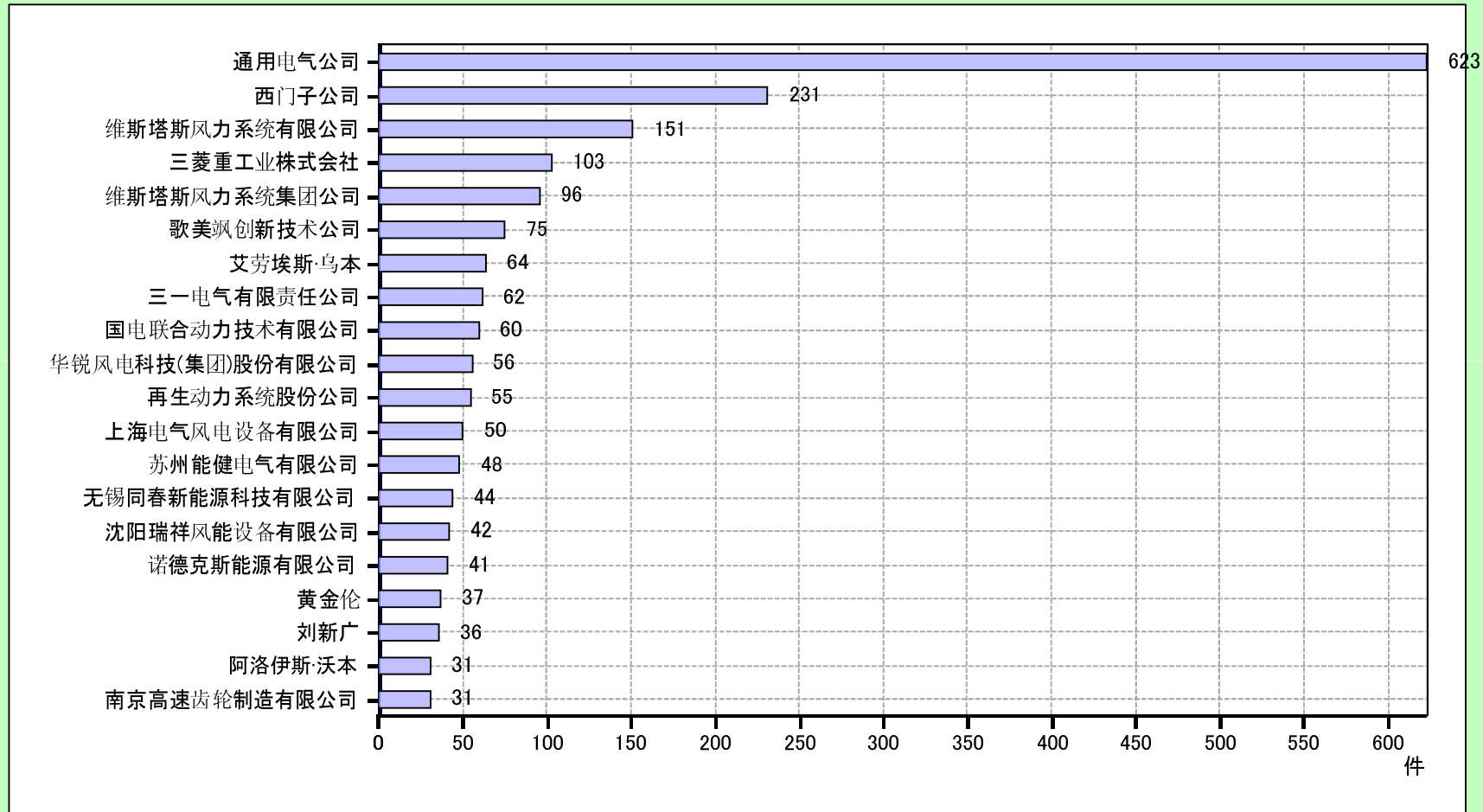
優先権主張国 * CN公報発行日(年)
件数マトリクスマップ

中国のみの**単独出願: 74.3%**
(OrbitのファミリーXPNを別途集計)



出願人別件数ランキング(中国語表記:非統制)

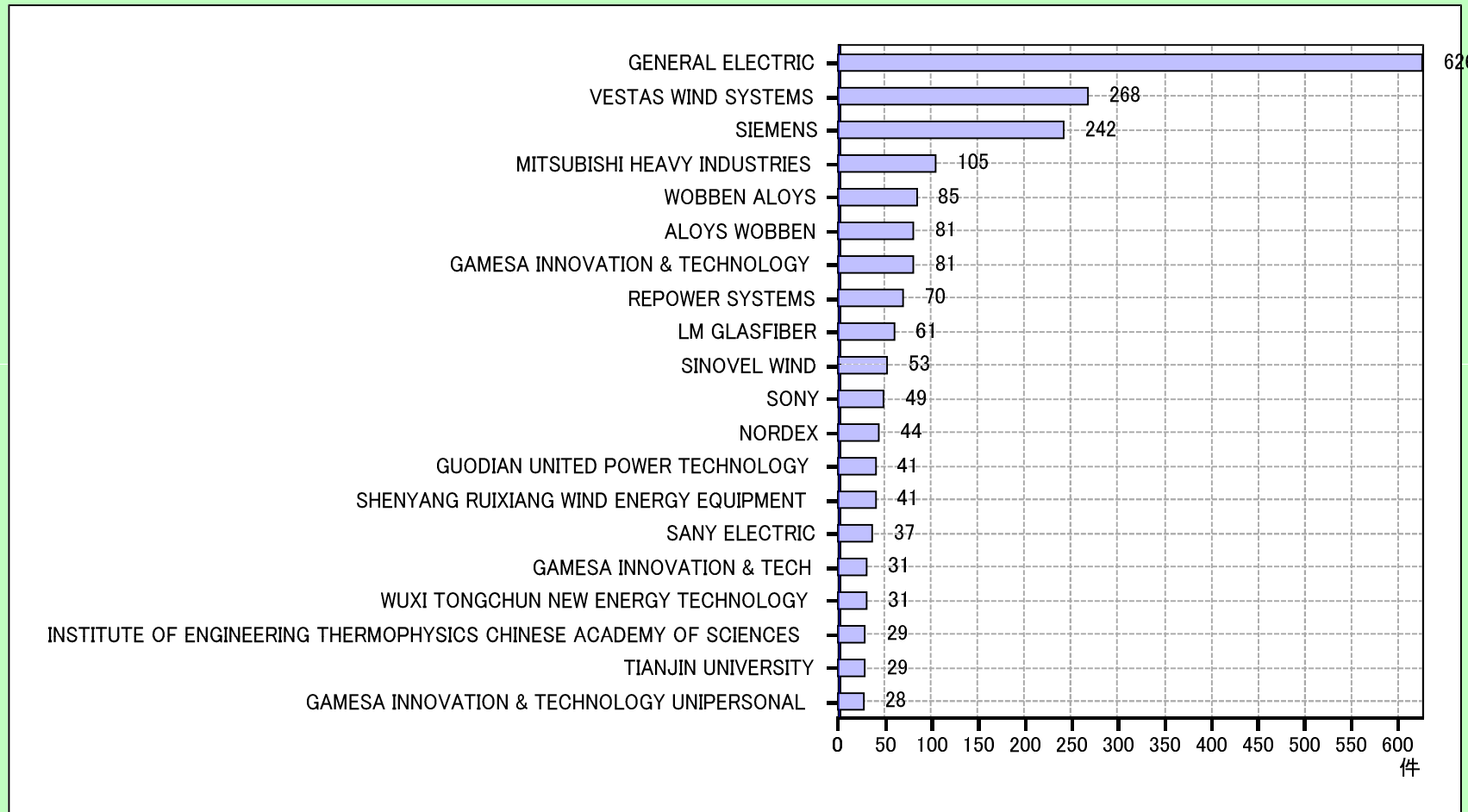
PAO出願人(CN)別
件数ランキングマップ



出願人別件数ランキング(英語表記:統制)

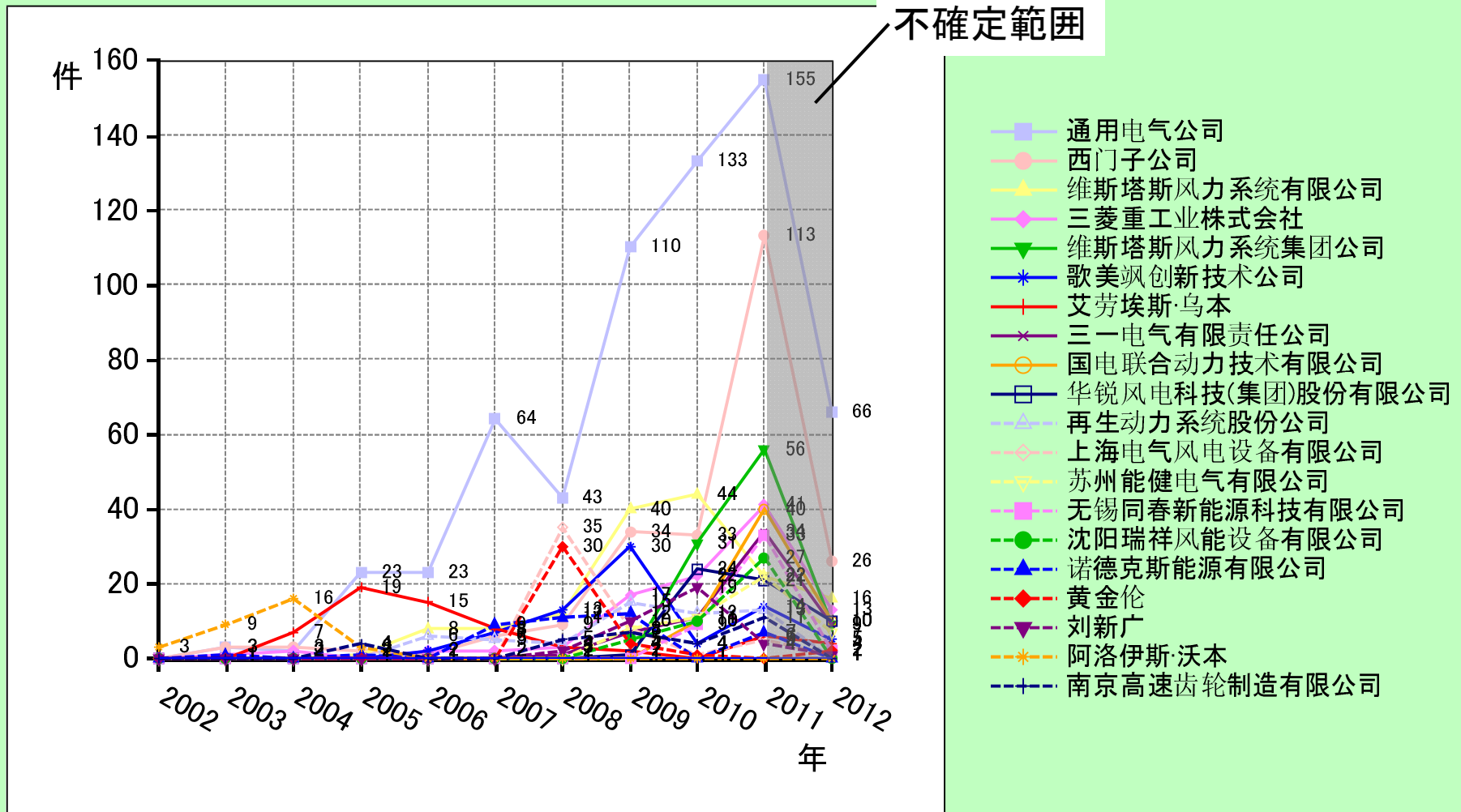
(Orbit 標準化出願人:英語)

出願人別 件数ランキングマップ



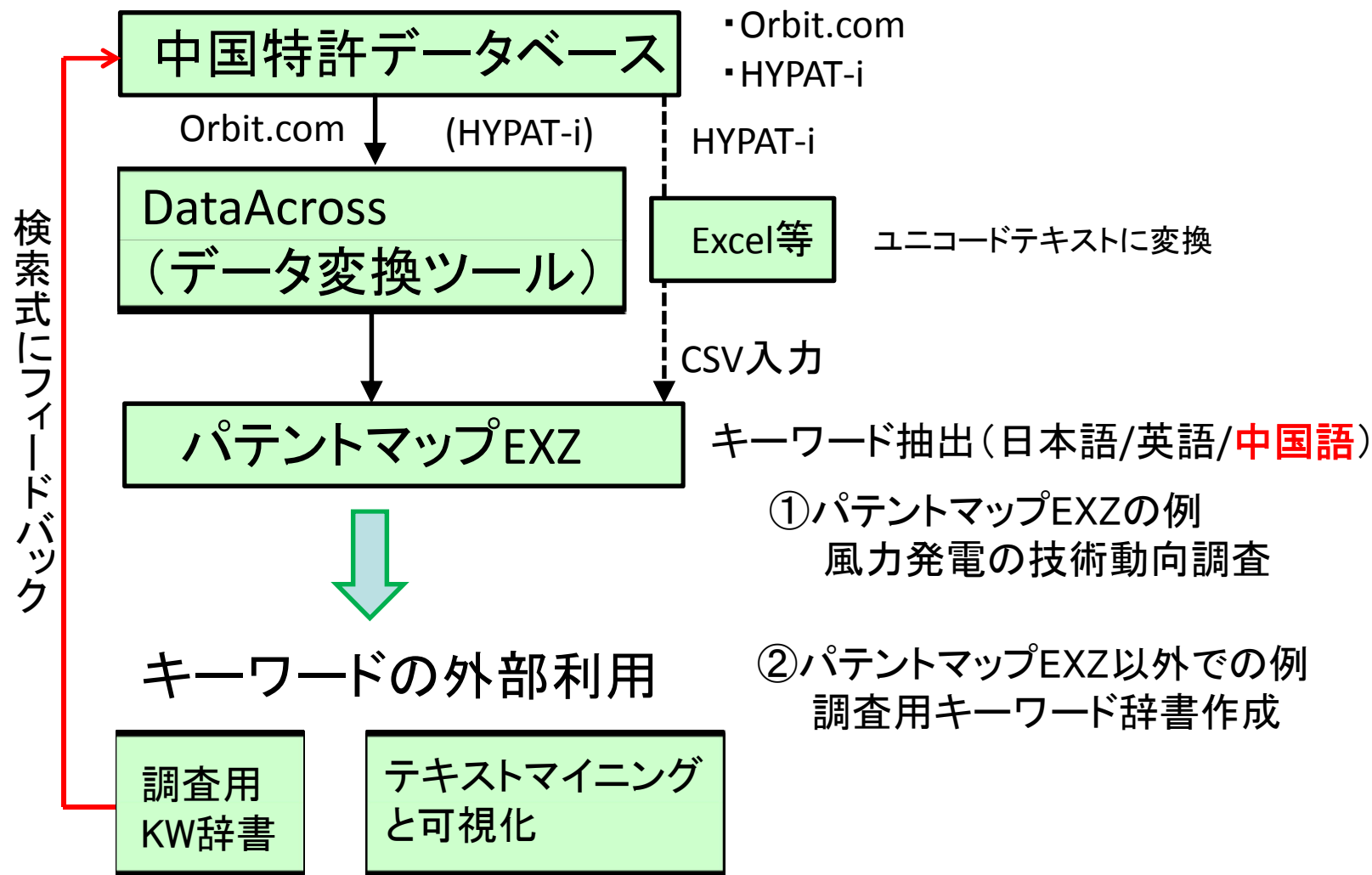
出願人別時系列マップ(中国語表記:非統制)

PAO出願人(CN)別
件数推移時系列マップ



3. パテントマップEXZによる中国語キーワード抽出

パテントマップEXZによるキーワード抽出フロー



中国語キーワード一覧

キーワード: キーワード一覧

部分一致 AND 実行 リセット

No.	データ名	すべて	発明の名称	要約	請求の範囲
1	権利	9654	0	7	9654
2	本	4199	0	4174	71
3	发电机	3204	72	2177	2888
4	叶片	2598	94	1729	2366
5	连接	2179	4	590	2088
6	风力发电机	1802	445	1482	1315
7	风	1678	43	1347	675
8	装置	1496	297	955	1046
9	固定	1438	3	512	1373
10	风能	1342	80	1193	385
11	轴承	1267	12	506	1213
12	转子	1157	30	695	1092
13	安装	1071	7	597	678
14	风轮	1025	35	794	853
15	轴	1010	11	437	954
16	控制	993	24	391	845
17	风力	933	45	739	399
18	旋转	925	0	424	764
19	转动	910	0	424	719
20	发电	890	23	674	393
21	电能	855	18	635	480
22	支架	812	1	456	777
23	风叶	793	18	612	720
24	风力涡轮机	789	279	577	731
25	塔架	769	18	509	686
26	两端	748	0	191	730
27	主轴	695	7	473	657
28	转轴	683	1	374	653
29	垂直	674	1	214	622
30	面	667	1	190	620

表示データ: 1~1000 (277457)



No.	データ名	すべて	発明の名称	要約	請求の範囲
1	権利	9654	0	7	9654
2	本	4199	0	4174	71
3	发电机	3204	72	2177	2888
4	叶片	2598	94	1729	2366
5	连接	2179	4	590	2088
6	风力发电机	1802	445	1482	1315
7	风	1678	43	1347	675
8	装置	1496	297	955	1046
9	固定	1438	3	512	1373
10	风能	1342	80	1193	385
11	轴承	1267	12	506	1213
12	转子	1157	30	695	1092
13	安装	1071	7	597	678
14	风轮	1025	35	794	853
15	轴	1010	11	437	954
16	控制	993	24	391	845
17	风力	933	45	739	399
18	旋转	925	0	424	764
19	转动	910	0	424	719
20	发电	890	23	674	393
21	电能	855	18	635	480
22	支架	812	1	456	777
23	风叶	793	18	612	720
24	风力涡轮机	789	279	577	731
25	塔架	769	18	509	686
26	两端	748	0	191	730
27	主轴	695	7	473	657
28	转轴	683	1	374	653
29	垂直	674	1	214	622

Excelへコピー&ペースト(1000ワード単位)、エクスポート(全ワード)⇒外部利用可能 13

中国語キーワード:部分一致 例「风」

キーワード:キーワード一覧

部分一致 AND 风 実行 リセット

No.	データ名	すべて	発明の名称	要約	請求の範囲
1	风力发电机	1802	445	1482	1315
2	风	1678	43	1347	675
3	风能	1342	80	1193	385
4	风轮	1025	35	794	853
5	风力	933	45	739	399
6	风叶	793	18	612	720
7	风力涡轮机	789	279	577	731
8	风速	622	5	359	365
9	风力发电装置	585	283	516	448
10	风力发电	505	54	447	154
11	风力发电机组	497	168	441	360
12	风力机	435	39	341	291
13	风轮机	356	115	248	315
14	风机	335	20	270	193
15	风车	313	47	250	219
16	风力发电设备	265	98	210	179
17	进风口	238	0	139	224
18	出风口	231	0	123	223
19	风力发电系统	228	93	190	158
20	风扇	190	2	99	177
21	迎风面	169	0	75	141
22	垂直轴风力发电机	165	79	137	128
23	风力涡轮机叶片	153	80	101	118
24	风轮叶片	135	15	88	104
25	风能发电	129	17	114	38
26	风向标	128	4	41	121
27	风轮轴	122	0	80	114
28	风力发电技术领域	121	0	121	0
29	风电机组	112	16	82	63
30	风电场	107	20	72	85

表示データ: 1~1000 (36883/277457)

4-1. 中国語キーワードを用いた特許情報解析

①パテントマップEXZの例 風力発電の技術動向調査

仮想キーワード件数ランキング

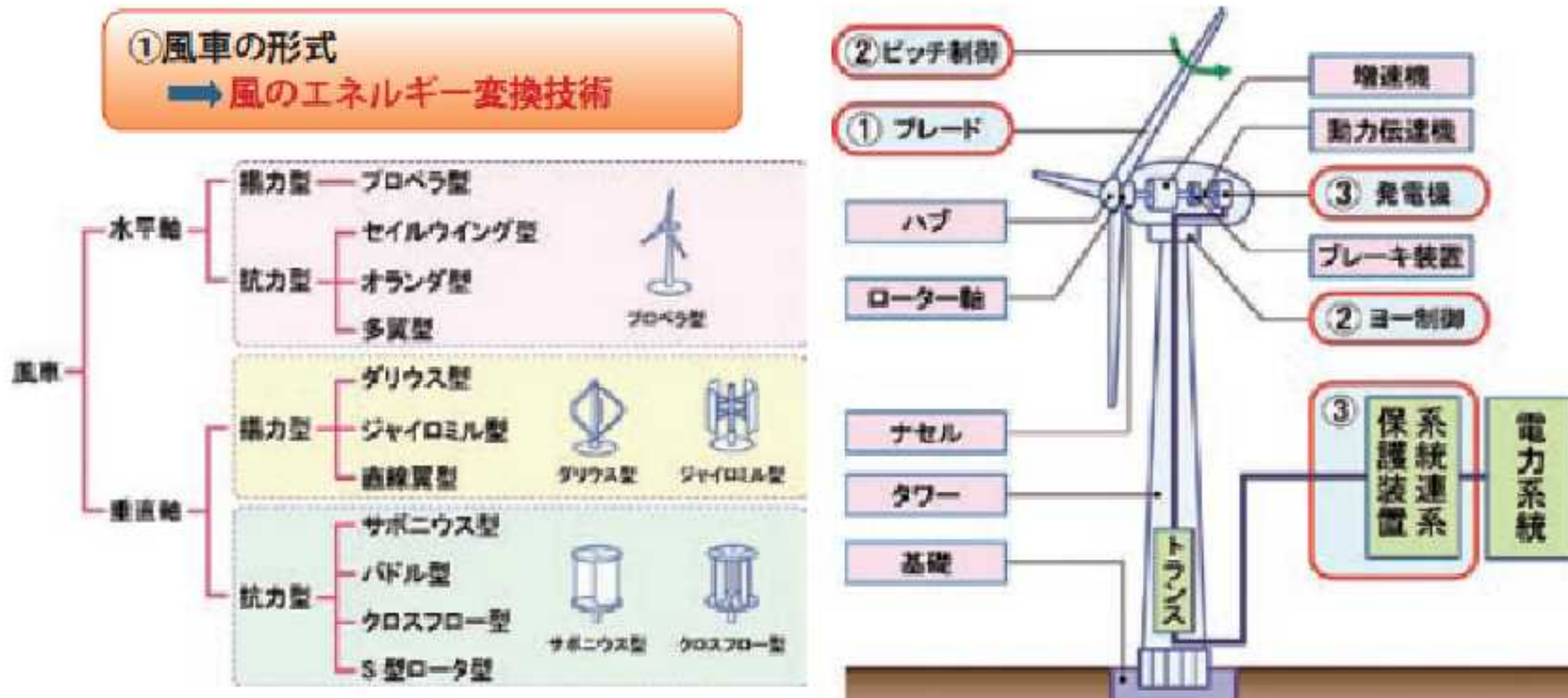
分類項目	要約キーワード	件数
制御方法	ピッチ(中文)	443
	ヨー(中文)	295
	ストール(中文)	47
制御目的	効率(中文)	1550
	強風(中文)	335
	弱風(中文)	242
	振動(中文)	204
制御項目	風向(中文)	579
	風速(中文)	1124
	回転速度(中文)	704
駆動方式	電気駆動(中文)	132
	油圧(中文)	304
	機械式(中文)	819
構成部品	ブレード(中文)	5417
	増速機(中文)	2029
	発電機(中文)	5896
	ブレーキ(中文)	475
	ナセル(中文)	570
	塔(中文)	1442
発電	発電(中文)	7215
洋上	洋上(中文)	106

制御方法の仮想項目

日本語	仮想項目	日本語	中国語
制御方法	ピッチ	ピッチ	间距
			螺距
			音高
			节距
			沥青
			俯仰
	ヨー・方位制御	ヨー	偏航
			横摆
		方位制御	定向控制
			計
ストール	ストール	失速	

仮想キーワードとはパテントマップEXZの仮想項目の一種で同義語・類義語を一つの仮想キーワードにまとめ(定義)たものである。例えば強風、暴風、台風を該当する中国語キーワードの同義語・類義語も含めて仮想キーワード:「強風」にまとめている。

風力発電関連技術(参考)

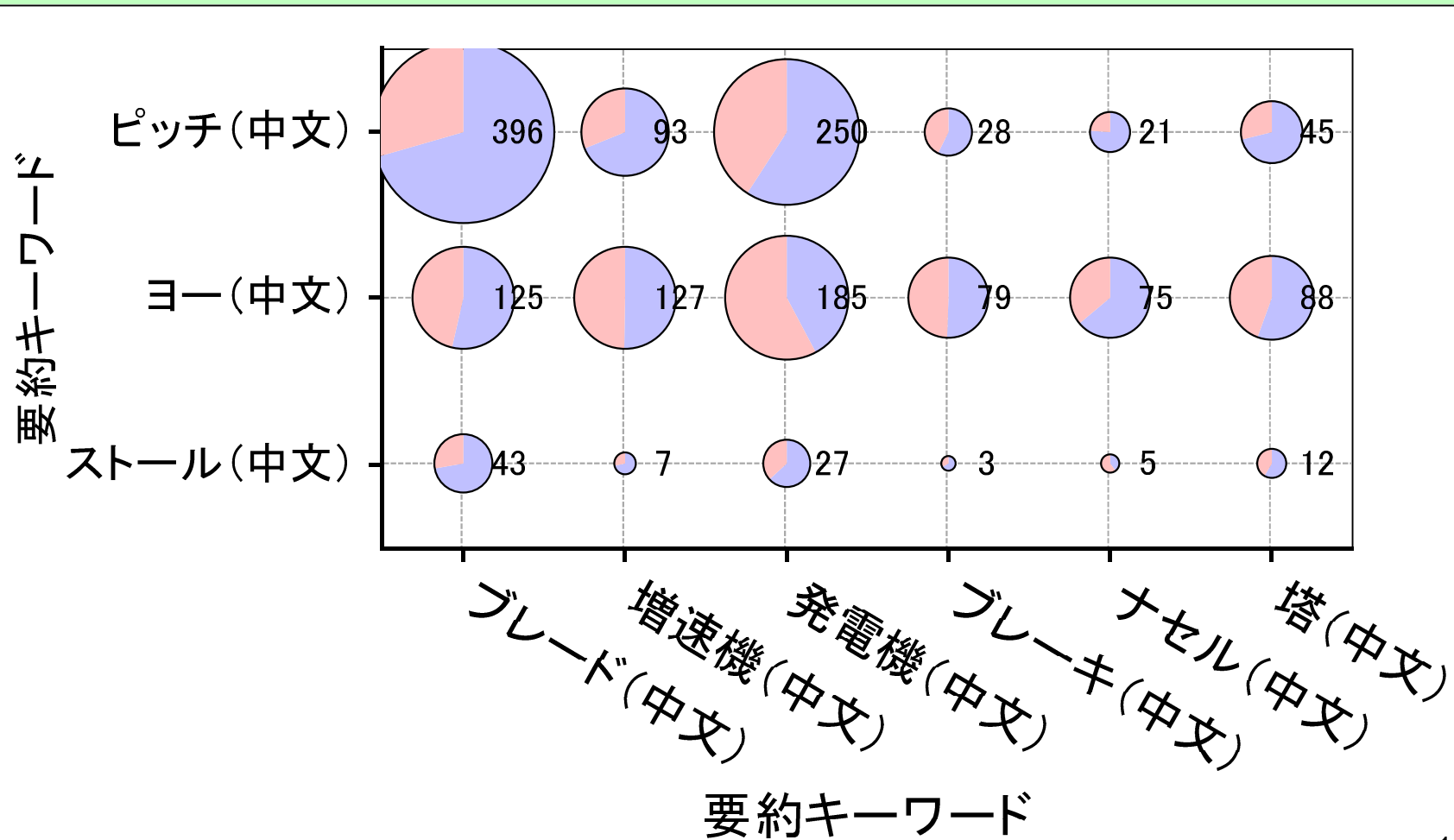


[5]平成22年度特許出願技術動向調査 風力発電
http://www.jpo.go.jp/shiryuu/pdf/gidou-houkoku/22wind_gene.pdf

制御方法KW * 構成部品 (要約)

件数マトリクスマップ

● 特許 ● 実用新案



4-2. 中国語キーワードを用いた特許情報解析

②パテントマップEXZ以外での例 調査用キーワード辞書作成

検索語	PATENTSCOPEの多言語検索CLIR			Google翻訳	
	EN	JA	ZH	日本語	英語
1 权利	rights	権利	权利	右	Right
2 本	local	ローカル	本	この	This
3 发电机	generator	発電	发电机	ジェネレータ	Generator
4 叶片	blade	羽根	叶片	ブレード	Blade
5 连接	connecting	連結	连接	接続	Connection
6 风力发电机	aerogenerator	風力発電機	风力发电机; 风力 发电机	風力タービン	Wind turbines
7 风	wind	風力	风	風	Wind
8 装置	device	装置	装置	機器	Equipment
9 固定	fixing	固定	固定	固定の	Fixed
10 风能	wind energy	風力エネルギー	风能	風力エネルギー	Wind energy
11 轴承	bearing	軸受	轴承	ベアリング	Bearing
12 转子	rotor	回転子	转子	ローター	Rotor
13 安装	mounting	取付	安装	インストール	Installation
14 风轮	windwheel	風力車輪 ウインドホイール	风轮; 风 轮	風力タービン	Wind turbine
15 轴	shaft	軸	軸	軸	Axis
16 控制	control	コントロール	控制	コントロール	Control
17 风力	wind power	風力発電	风力	風力	Wind power
18 旋转	rotating	回転	旋转	回転	Rotation
19 转动	rotary	回転式	转动	ターン	Turn
20 发电	power generation	発電	发电	電気を発生させる	Generate electricity
21 电能	electric energy	電力量	电能	電気	Electricity
22 支架	support	支持(体)	支架	サポート	Support
23 风叶	fan blade	ファンブレード	风叶; 风 叶	ブレード	Blades
24 风力涡轮机	wind turbine	風車	风力涡轮机	風力タービン	Wind turbines
25 塔架	tower	タワー	塔架; 塔 架	タワー	Tower

多言語検索CLIR (Cross-Lingual Information Retrieval)

パテントマップEXZで抽出した中国語キーワード

WIPO PATENTSCOPEの多言語検索CLIR

中国語の「**风力发电机**」の検索例

Deutsch | English | Español | Français | 日本語 | 한국어 | Português | Русский | 中文 |

WIPO PATENTSCOPE
国内特許及びPCTコレクションを検索:CLIR

Home > IP Services > PATENTSCOPE > Back to PATENTSCOPE

検索用語を入力

検索語 [Help]

风力发电机

» 検索言語: 中国語

» 拡張モード: 自動
Automatic

» 正確性を高める (適合率) [Slider] 網羅性を高める (再現率)

検索

注) 通常の検索では「設定 (Supervised)」の方が細かな指定が可能。

「多言語検索CLIR」の仕組み自体が非常に参考になる

WIPO PATENTSCOPEの多言語検索CLIR

WIPO PATENTSCOPE
国内特許及びPCTコレクションの検索

WORLD INTELLECTUAL PROPERTY ORGANIZATION

検索 | 閲覧 | 翻訳 | オプション | News | ログイン | ヘルプ

Home > IP Services > PATENTSCOPE

検索結果1-10 of 1,808 for 検索基準: (ZH_Tl:("风力发电机" OR "风力 发电机"~21) OR ZH_AB:("风力发电机" OR "风力 发电机"~21)) OR (DE_Tl:("Aerogenerator") OR DE_AB:("Aerogenerator")) OR (EN_Tl:("aerogenerator") OR EN_AB:("aerogenerator")) OR (ES_Tl:("aerogenerator") OR ES_AB:("aerogenerator")) OR (FR_Tl:("aérogénérateur") OR FR_AB:("aérogénérateur")) OR (JA_Tl:("風力発電機") OR JA_AB:("風力発電機")) OR (KO_Tl:("풍력 발전기") OR KO_AB:("풍력 발전기")) OR (PT_Tl:("gerador de força energia eólica"~22) OR PT_AB:("gerador de força energia eólica"~22)) OR (RU_Tl:("аэрогенератор") OR RU_AB:("аэрогенератор")) 特許庁/PCT: all言語: JA語幹処理:true

戻る 1 2 3 4 5 6 7 8 9 10 次へ Page:1 / 181 Go >

絞り込み検索 検索 RSS Query Tree

検索結果の分析

検索クエリーより各国語Σを抽出

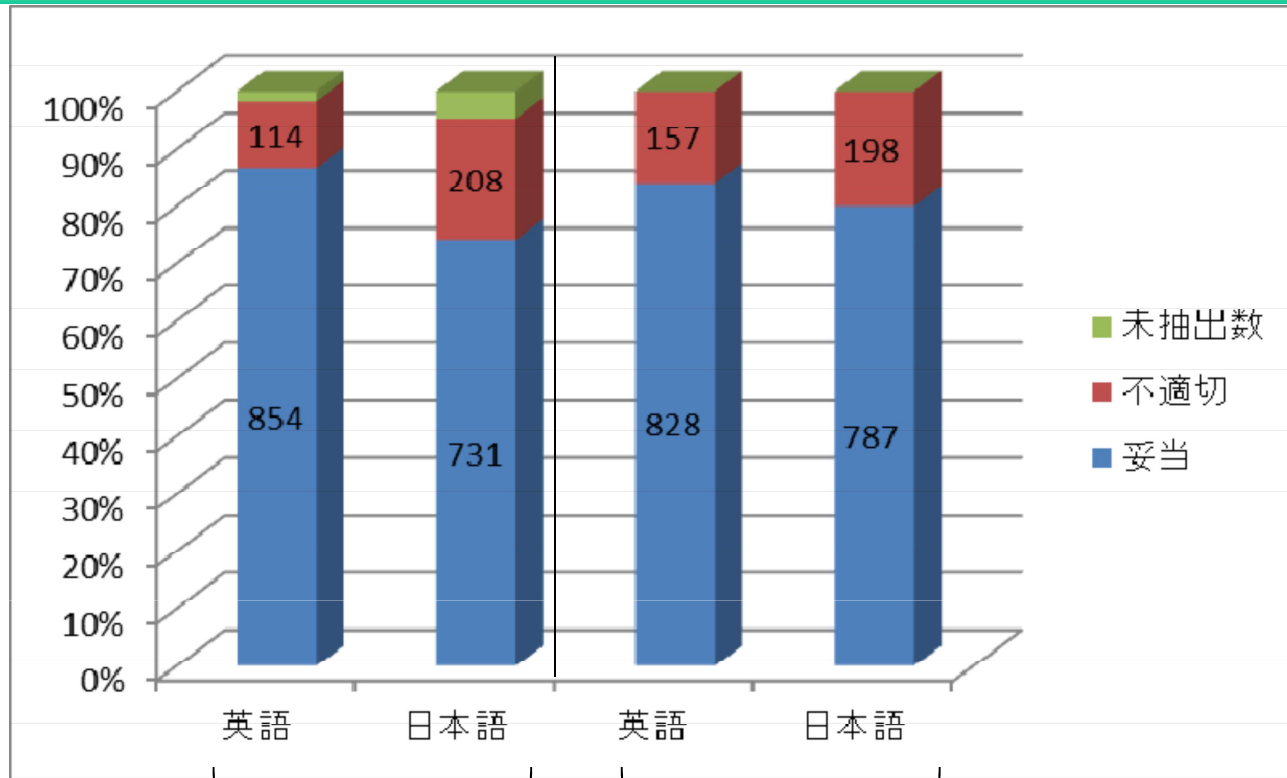


Excel上に風力発電関連の英語、ドイツ、スペイン、フランス、日本、韓国、ポルトガル、ロシア、中国語の多言語辞書を構築

検索語	検索言語	拡張モード	適合率/再現率	EN	DE	ES	FR	JA	KO
风力发电机	中国語	自動	1: 正解性を高める	aerogenerator	Aerogenerator	aerogenerator	aérogénérateur	風力発電機	풍력 발전기

PT	RU	ZH
gerador de força energia eólica	аэрогенератор	风力发电机; 风力 发电机

キーワード辞書の妥当性検討



	CLIR		Google 翻訳	
	英語	日本語	英語	日本語
抽出語数	968	939	985	985
未抽出数	17	46	0	0
妥当	854	731	828	787
不適切	114	208	157	198

妥当性は意味内容から人が判断、英語はCLIR、日本語はGoogle翻訳が良い。CLIR、Google翻訳併用で精度向上見込。

5. 解析結果の中国特許調査への応用

統計的解析

調査用キーワードの検討例

英語の"pitch"に相当する中国語キーワードは？

検索語	EXZ: すべて		Orbit 件数	CNIPR 件数	CLIR 件数	本法: 要約	
	順位	件数				DF	TF
pitch			629		767688		
间距	228	179	105	114	1209867	108	152
螺距	3629	14	23	22	635671	20	65
桨距	1034	46	277	293	510750	264	637
桨距角	408	108	75	94	12912	76	135
俯仰	3911	13	41	35	115187	34	69
俯仰角	3146	16	14	13	600967	13	21
节距	5319	10	26	23	644408	20	45
沥青	24551	3	0	0	575275	0	0
音高	83012	1	0	1	321926	1	1

ノイズ？

ランキング
降順

表題 要約
要約

DF: 文書頻度
TF: ターム頻度

意味的解析 解析結果の中国特許調査への応用

多言語検索CLIR (Cross-Lingual Information Retrieval)

検索語	検索言語	拡張モード	適合率/再現率	EN	JA	ZH
间距	中国語	自動	1: 正解性を高める	interval	間隔	间距
螺距	中国語	自動	1: 正解性を高める	pitch	ピッチ	螺距
桨距	中国語	自動	1: 正解性を高める	pitch	ピッチ	桨距
桨距角	中国語	自動	1: 正解性を高める	pitch angular	ピッチ アンギュラ	桨距 角
俯仰	中国語	自動	1: 正解性を高める	pitching	型ピッチング	俯仰
俯仰角	中国語	自動	1: 正解性を高める	pitch angle	型ピッチング アンギュラ	俯仰角;俯仰 角
节距	中国語	自動	1: 正解性を高める	pitch	ピッチ	节距
沥青	中国語	自動	1: 正解性を高める	asphalt	アスファルト	沥青
音高	中国語	自動	1: 正解性を高める	pitch	ピッチ	音高
间距	中国語	自動	2	interval;distance;gap;spacing;pitch;clearance;space length	間隔;音程;間隙;区間;隙間;ギャップ;インターバル	间距;间隙
螺距	中国語	自動	2	pitch	ピッチ;型ピッチング	螺距;俯仰
桨距	中国語	自動	2	pitch	ピッチ	桨距;俯仰
桨距角	中国語	自動	2	pitch angle;inclination angle;tilt angle	ブレードピッチ角制御;のピッチ角;ブレードピッチ角	桨距角;桨距 角;傾斜角;俯仰角;桨距 位置;俯仰 角;俯仰 位置
俯仰	中国語	自動	2	tone;pitching;tonal;pitch variable;luffing;cabrage;variable amplitude;amplitude	トーン;階調;投球;建込み;型ピッチング;ピッチ運動;張り;ピッチング制御;ピッチ運動制御	俯仰;音调;螺距;用动;变幅
俯仰角	中国語	自動	2	asphalt;pitch;pitch angle;inclination angle;tilt angle	ピッチ;アスファルト;ブレードピッチ角制御;のピッチ角;ブレードピッチ角	俯仰角;俯仰 角;节距;沥青;桨距;傾斜角;俯仰 斜;俯仰 交錯;节距 角
节距	中国語	自動	2	asphalt;bituminous;bitumen;pitch	ピッチ;アスファルト	节距;进行俯仰;沥青
沥青	中国語	自動	2	asphalt;bituminous;pitch;bitumen;asphaltum;pitch variable;guide	アスファルト;ピッチ;ピッチ可変;ピッチ可変式	沥青;导程;螺距
音高	中国語	自動	2	pitch;fundamental tone	ピッチ;音高	音高;基音
pitch	英語	自動	1: 正解性を高める	pitch	ピッチ	桨距
pitch	英語	自動	2	pitch;height	ピッチ;向;高度;高低;揚高;高検知;身長;車高;程	音调;桨距;基音;螺距;沥青;导程;俯仰;音高;节距

ノイズ?

意味的解析 解析結果の中国特許調査への応用

抽出 中国語	北辞郎(中日辞書)	CLIR		Google翻訳	
		日本語	英語	日本語	英語
间距	1.距離. 間隔 2.<電脳>間隔. ピッチ	間隔	interval	間隔	Spacing
螺距	1.(ねじの)ピッチ 2.(巻いているワイヤなどの)ピッチ	ピッチ	pitch	ピッチ	Pitch
桨距		ピッチ	pitch	ピッチ	Pitch
桨距角		ピッチ アンギュラ	pitch angular	ピッチ角	Pitch angle
俯仰	1.下を向くことと上を向くこと 2.一挙一動	型ピッチング	pitching	ピッチ	Pitch
俯仰角	ピッチ角	型ピッチング アンギュラ	pitch angle	ピッチ角	Pitch angle
节距	ピッチ(pitch)	ピッチ	pitch	ピッチ	Pitch
沥青	1.瀝青(以下のものを含む総称) 2.コールター ル(原料は石炭) 3.アスファルト(原料は石油)	アスファルト	asphalt	アスファルト	Asphalt
音高	<音>音高. ピッチ	ピッチ	pitch	ピッチ	Pitch

ノイズ

北辞郎(中日辞書)

<http://www.ctrans.org/>

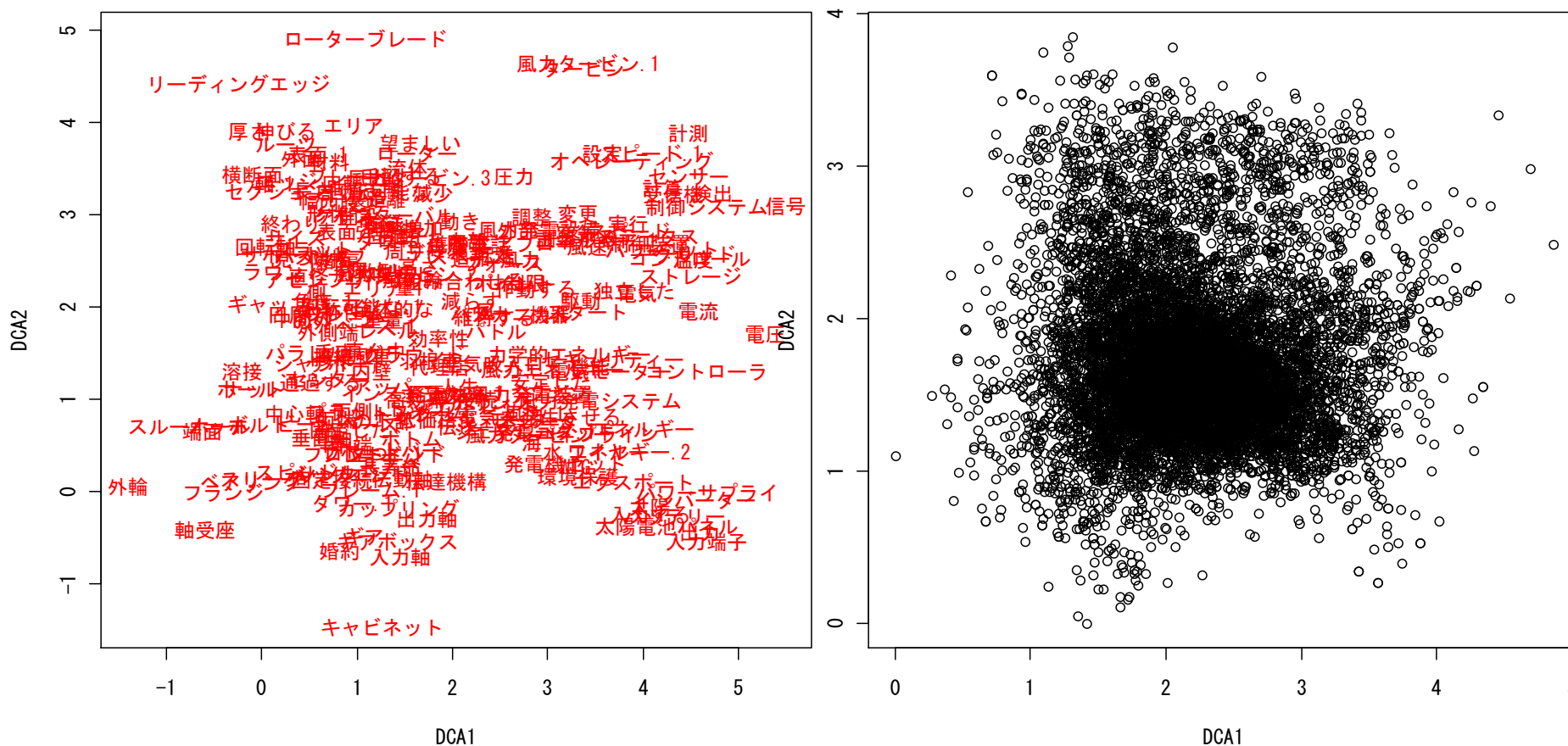
中国特許公報より抽出した中国語キーワードを基に作成したユーザー辞書を用いて適切なキーワードを選択することで中国特許調査の調査精度、再現率の向上に有用である。用語の選択が重要である。

統計的解析 中国語KWの統計的解析(対応分析)

中国語KWによるテキストマイニングと可視化

中国語250語相対的ポジション
(表示はGoogle翻訳)

10480件の公報の相対的ポジション

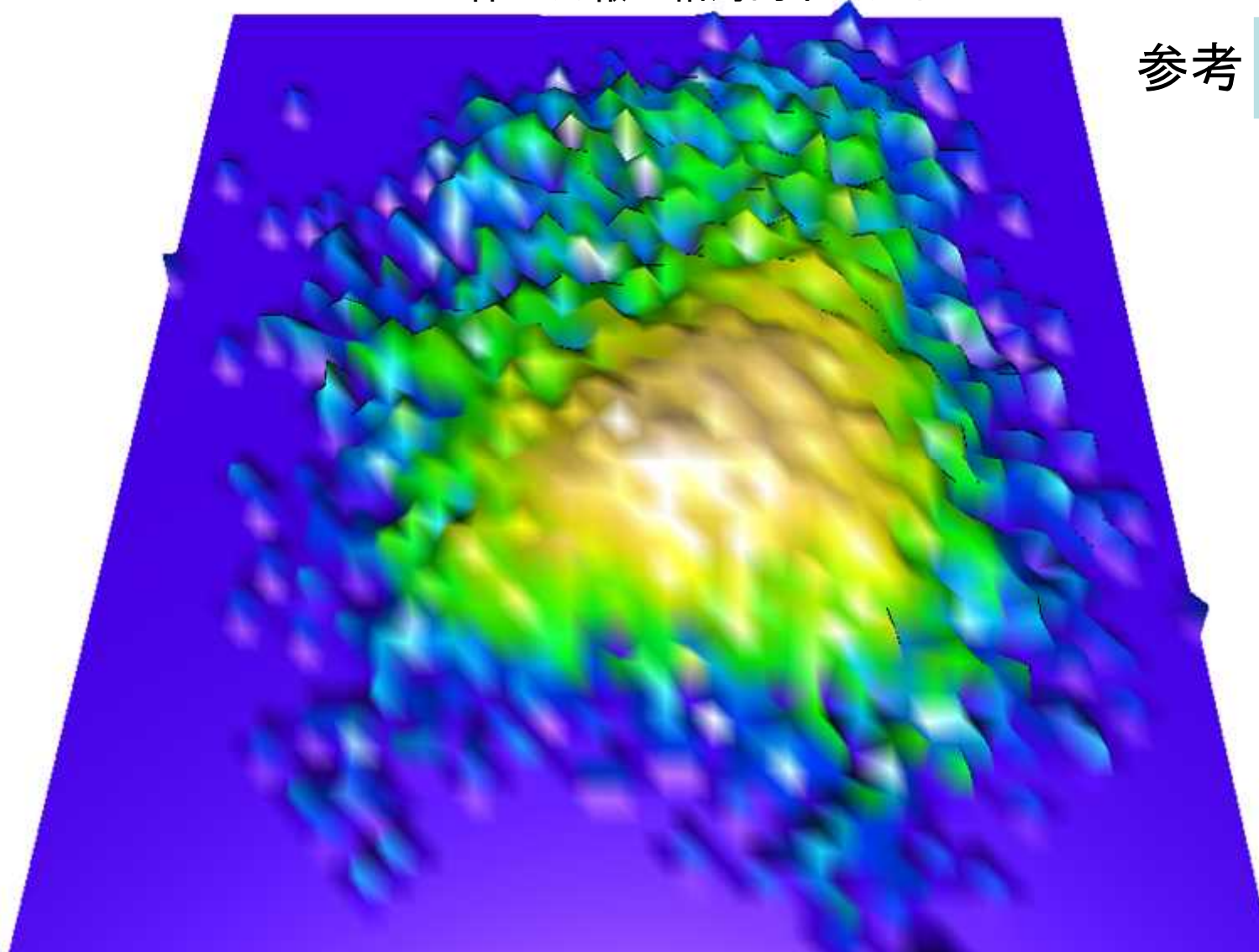


使用データ: HYPAT-i タイトル、要約、請求項1の中国語ダウンロードデータ

統計解析言語Rおよびpackage:ade4, vegan使用

10480件の公報の相対的ポジション

参考



rgl.surface関数による3D表示

6. まとめ

パテントマップEXZにより抽出した中国語キーワードを用いた特許情報解析に関して、下記の検討を行った。

- ①中国語キーワード抽出
- ②特許情報解析
- ③英語キーワード解析結果との比較
- ④解析結果の中国特許調査への応用

中国特許公報より抽出した中国語キーワードを基に作成したユーザー辞書を用いて適切なキーワードを選択することで中国特許調査の調査精度、再現率の向上に有用である。用語の選択が重要である。

中国語KWによるテキストマイニングと可視化も可能。

謝辞

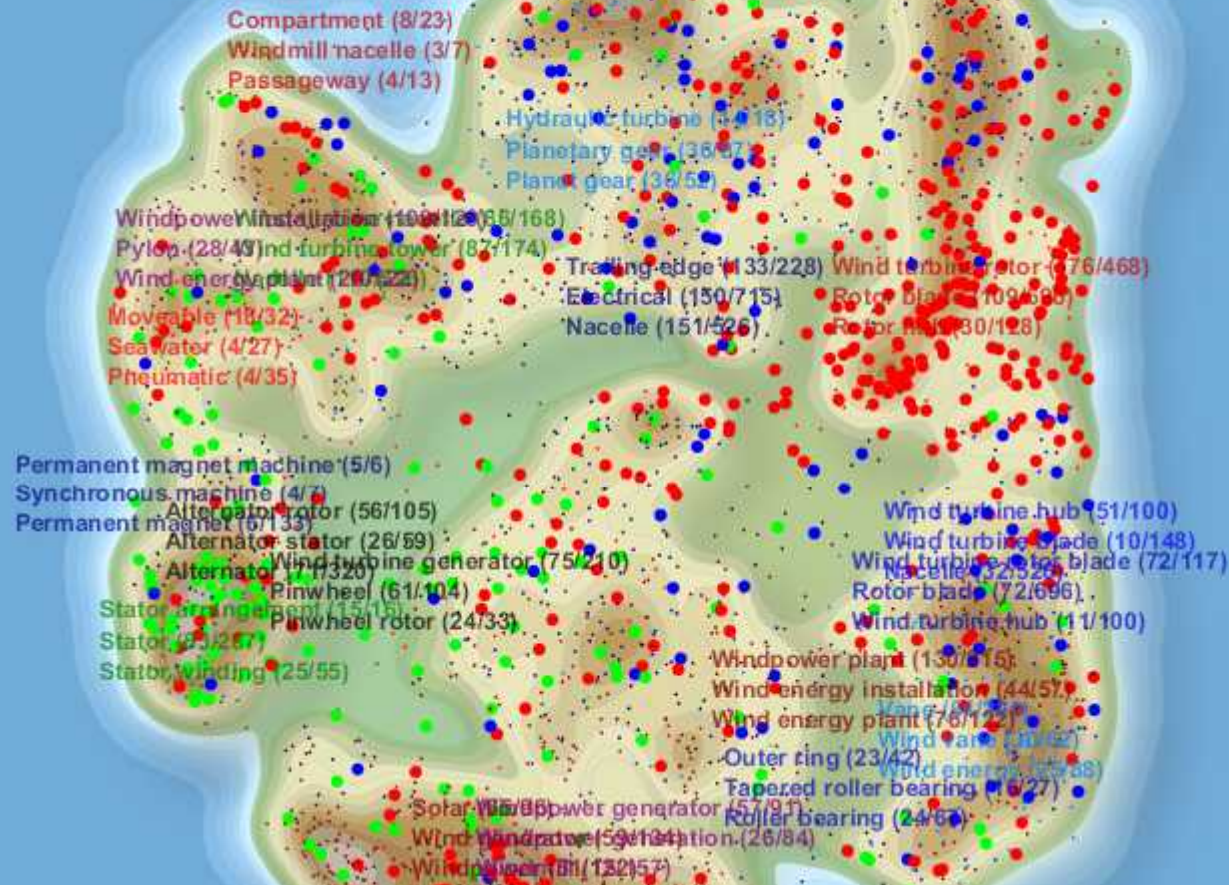
「謝辞」

本報告は2012年度の「アジア特許情報研究会」のワーキングの一環として報告するものであり、報告者として名前を挙げさせていただいた他に、オリンパスメディカルシステムズの中西昌弘氏にはCLIRの多言語キーワード抽出で、神戸製鋼所の石田政司氏には抽出した中国語キーワードの検討で多大な協力をしていただきました。ここに改めて感謝申し上げます。

中国特許情報解析チーム一同

ご清聴、ありがとうございました。

参考2: 英語キーワードによる解析



Orbit Analysis module (beta版)
Landscape Map

	WIND TURBINE ROTOR	
	WIND TURBINE GENERATOR	
	WIND TURBINE BLADE	

参考3 「設定 (Supervised)」モードの分野別辞書



Deutsch | English | Español | Français | 日本語 | 한국어 | Português | Русский | 中文 |

国内特許及びPCTコレクションを検索:CLIR

Home > IP Services > PATENTSCOPE > Back to PATENTSCOPE



検索用語を入力

検索語

技術分野 [MECH]

[Help]

[ADMN] 管理、ビジネス、経営及び社会科学
[AERO] 航空宇宙工学
[AGRI] 農業、漁業及び林業
[AUDV] オーディオ、オーディオビジュアル、画像及びビデオ
[AUTO] 自動車及び道路車両工学
[BLDG] 土木及び建築工学
[CHEM] 化学及び材料技術
[DATA] コンピュータ科学及び通信放送
[ELEC] 電気工学及びエレクトロニクス
[ENGY] エネルギー、燃料及び伝熱工学
[ENVR] 環境及び安全工学
[FOOD] 食品及び食品技術
[GENR] 一般性、言語及びメディア情報科学
[HOME] 家財の内容及び家庭のメンテナンス
[HORO] 精密機械、宝石及び時計
[MANU] 製造及び材料処理技術
[MARI] 海洋技術
[MEAS] 標準化、単位、計量及び試験
[MEDI] 医療技術
[METL] 金属工学
[MLI] 軍事技術
[MINE] 鉱業、石油ガス抽出及び鉱物
[NANO] ナノテクノロジー
[PACK] 包装及び商品の流通
[PRNT] 印刷及び紙
[RAIL] 鉄道工学
[SCIE] 光工学
[SPRT] スポーツ、レジャー、観光及びサービス業
[TEXT] 繊維衣料産業
[TRAN] 交通

追加

削除

[MECH] 機械工学