

# A22

## 英語・原語によるハイブリッド検索： PatBase, QPAT(Orbit.com), Discover による 英語・原語ハイブリッド検索の検討

○富士フイルム株式会社	田畑文也
株式会社 神戸製鋼所	石田政司
株式会社 アイピックス	水町保宏

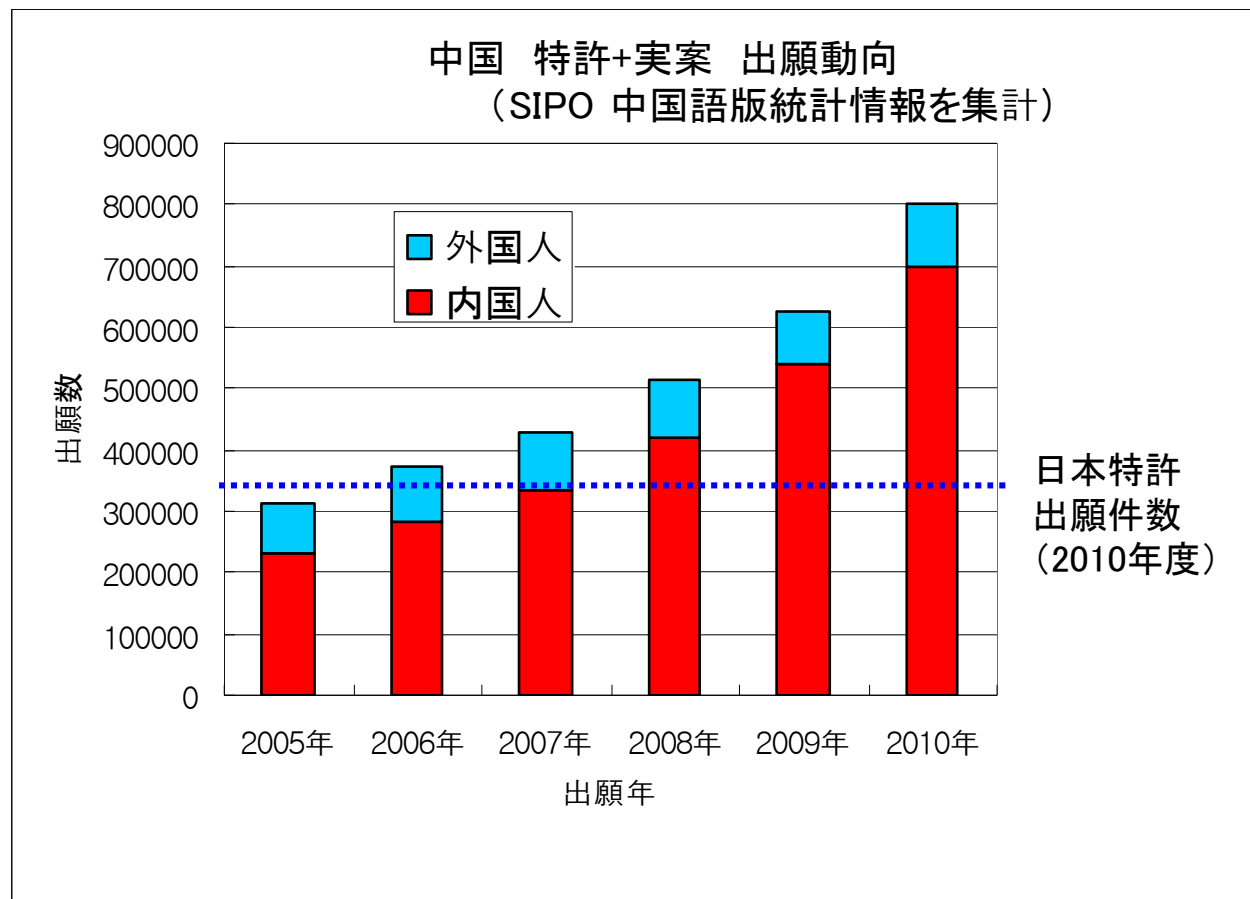
# 発表内容

1. はじめに
2. 中国特許調査法とDB
3. 中国特許収録確認
4. 出願人検索検証
5. キーワード検索検証
6. まとめ

(本報告内容は、アジア特許情報研究会  
2011年度ワーキング活動の一部を報告)

# 1.はじめに

## 1.1 中国特許(特許+実案)出願件数推移

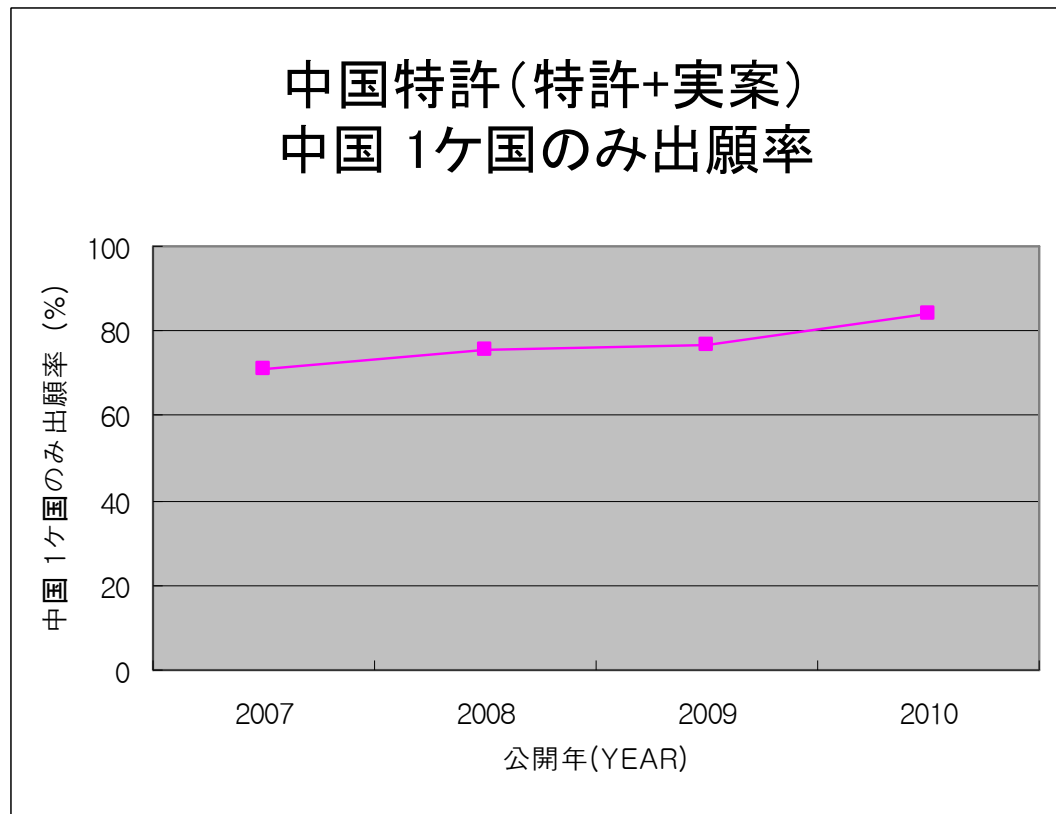


中国特許(特許+実案)の出願数は急激に増大し、かつ内国人出願が多い。

図1 中国 特許+実案 出願動向

# 1.はじめに

## 1.2 中国特許(特許+実案) 中国1ヶ国のみ出願率動向



中国特許(特許+実案)は、中国から海外への出願が少なく、対応ファミリー特許の人間翻訳英語全文等による情報が限られる。

(検索式例) STN WPI  
S ((CNA? or CNY? or CNU?)/PK)(P)(2008/PY)  
S L1 and 1/CYC

図 2 中国特許(特許+実案) 1ヶ国出願率

## 2.中国特許調査法とDB

### 2.1 網羅性を上げる従来の手法例

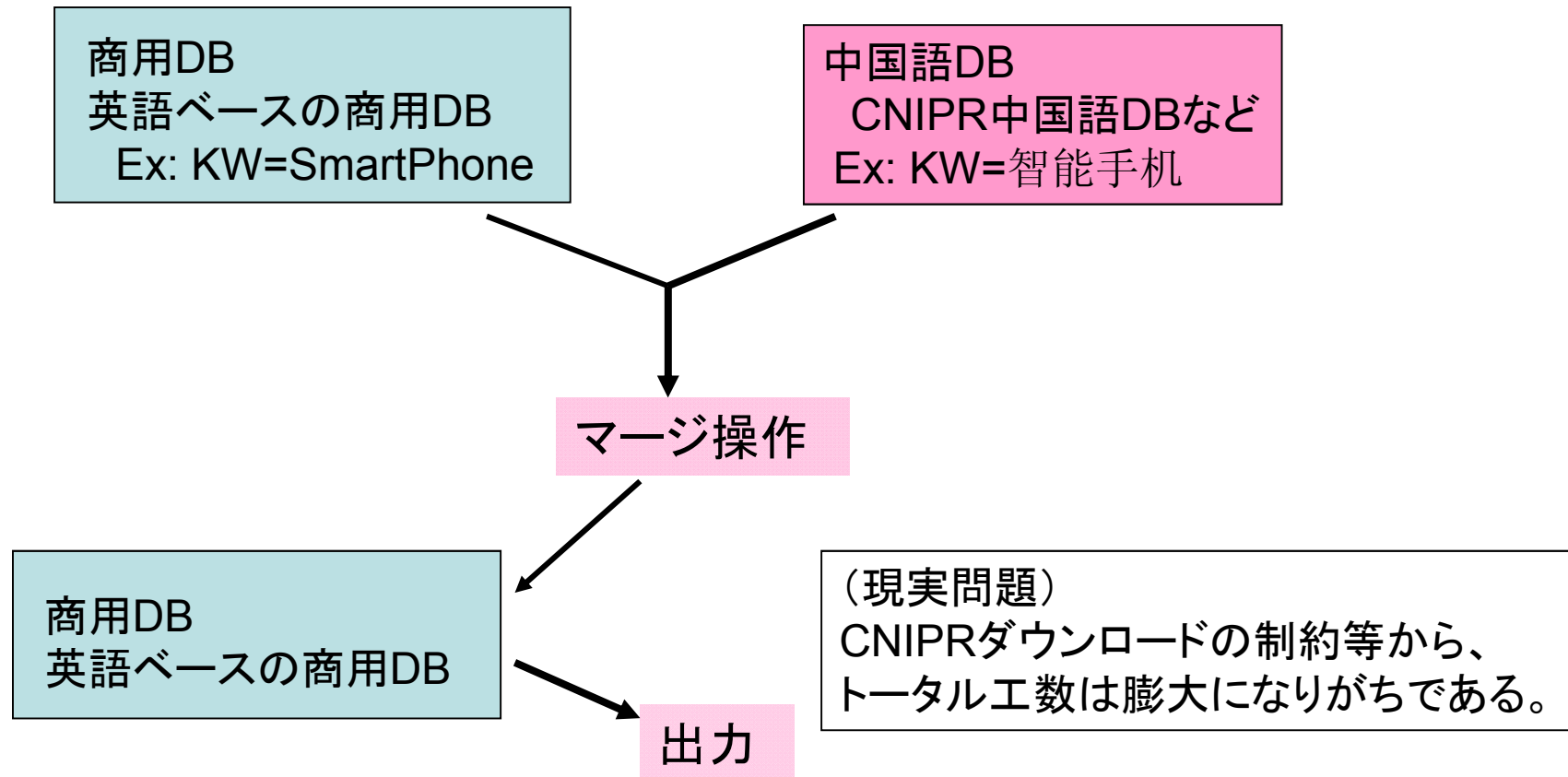


図 3 複数DBを用いた際のスキーム例

## 2.中国特許調査法とDB

### 2.2 中国特許DBとデータソース

クレーム以下のデータは、**IPPH(中国国家知識産権局 知識産権出版社)**から、中国語及び機械翻訳の英語データを購入するか、独自人間翻訳、独自OCR処理などをデータソースとしている。 →それぞれにメリット、デメリット有り

表 1 中国特許DB例とクレーム以下のデータソース

DB	クレーム	全文	中国語検索
ThomsonInnovation	2007~ 独自人間翻訳	収録無し	×
PatBase	独自OCR-機械翻訳	独自OCR-機械翻訳	○
TotalPatent	IPPH-機械翻訳	IPPH-機械翻訳	×
CPA Global Discover	IPPH-機械翻訳	IPPH-機械翻訳	○
Orbit.com (QPAT)	IPPH-機械翻訳	IPPH-機械翻訳	○
ShareResearch	IPPH-機械翻訳	IPPH-機械翻訳	×
DialogFile325	独自OCR-機械翻訳	独自OCR-機械翻訳	×

## 2. 中国特許調査法とDB

### 2.3 英語と原語(中国語)のハイブリッド検索

- いくつかの商用DBにて、出願人検索、キーワード検索などで、英語だけでなく、原語(中国語等)で検索できる機能が付与されてきた。
- 2011年7月時点で日本に上市されているファミリータイプの商用DBで、**非ラテン語系の多言語検索**に対応しているDBは以下の3つである。

- ① PatBase (RWSグループ社)
- ② Orbit.com、QPAT (Questel社)
- ③ Discover (CPA社)

今回はこの3つのDBでの、多言語検索(以下、**ハイブリッド検索**)について発表する。(今回は中国語のみ検討)

(評価用IDは、それぞれ(株)RWSグループ、(株)プロパティ、日本技術貿易(株)より、提供いただきました。(感謝) )

# 2.中国特許調査法とDB

## 2.4 ハイブリッド検索画面例 (Orbit.com)

Orbit.com

キーワード

コンテンツDB: FAMPAT (ファミリータイプDB)

中国特許の英語コンテンツ内を検索

中国特許の中国語コンテンツ内を検索

一つのDB内で、英語、中国語の結果を自動マージ処理

Orbit.com  
QPAT  
国内代理店  
(株)プロパティ  
(株)パトリス

名称	オリジナル
(A) One kind has both main from equipment's smartphone and	
(U) 一种利用智能手机控制家用电器的系统	
(A) More than one kind of operating system smartphones [Ma	
(A) Automatic Switcher features a method of achieving mobile	
(U) The interaction between a smart phone GPS [Machine Tran	
(U) Smartphone outside sets at the battery [Machine Translati	

図 4 Orbit.comでのハイブリッド検索例



# 2. 中国特許調査法とDB

## 2.5 ハイブリッド検索結果表示例(PatBase)

検索結果は、英語で詳細確認可能  
(クレーム以下も)



CN201365348 Y [Translate to English](#) (via PatBase internal translation) or Select language to translate to (via Google):

English

Title :

一种无线上网的智能手机

A wireless Internet smartphones

Abstract :

一种无线上网的智能手机, 包括一手机本体3, 在手机本体3上设置有USB接口7, USB接口7上的USB主控芯片2的D0-D7共8条数据信号线与微处理器1的DQ0-DQ7共8条数据信号线一一对应相连, USB接口7上连接有无线网卡4, 无线网卡4将指令转化为无线信号发出, 由AP5接收并传入Internet线路6进行数据交换, 无线AP5再将信号转化为无线的信号发射出去, 由无线网卡4接收, 从而实现智能手机无线上网, 具有费用低、速度快的特点。

A wireless Internet smart phones including a mobile body 3, the body 3 is set on the phone with USB interface 7, USB interface, the USB controller chips, 7 2 8 D0-D7 of the data signal line and the microprocessor 1 of 8 DQ0-DQ7 data signal line connected to one correspondence, USB interface 7 is connected to the wireless card 4, the wireless card 4 into the wireless signal instructions issued by the AP5 receive and incoming Internet data exchange line 6, AP5 into the wireless signal and then transmitted out the wireless signal, received by the wireless LAN 4, wireless Internet access in order to achieve the smart phone with low cost and fast speed.

Claim(s) :

1、一种无线上网的智能手机, 包括一手机本体(3), 其特征在于, 在于机本体(3)上设置有USB接口(7), USB接口(7)上的USB主控芯片(2)的D0~D7共8条数据信号线与微处理器(1)的DQ0~DQ7共8条数据信号线一一对应相连, USB接口(7)上连接有无线网卡(4), 无线网卡(4)将指令转化为无线信号发出, 由AP5接收并传入Internet线路(6)进行数据交换, 无线AP5再将信号转化为无线的信号发射出去, 由无线网卡(4)接收, 从而实现智能手机无线上网, 具有费用低、速度快的特点。

1, a wireless Internet smart phones including a mobile body (3), characterized in that, that machine body (3) plus or minus to set a USB connection P (7). USB connection P (7) plus or minus the USB

PatBase  
国内代理店  
(株)RWSグループ

図 5 PatBaseでの検索結果表示例 (閲覧モード)

# 2. 中国特許調査法とDB

## 2.6 原語キーワード取得方法例

Welcome

CPA GLOBAL DISCOVER

Home Search

Query String

Keywords Publication Number Similarity

Type or paste keywords below

TI:("smartphone" or "smart phone")PC("CN")

CN 201234298 Y: Expanding apparatus for smart phone  
Publication Date: 2009-05-06

Return to results Find Similar Exclude Similar

Pane Layout  Keyword Highlighting Select View My Custom Go Select Action Go 21 Result 1 of

View in this pane: Bibliography

View in this pane: Other Languages

Publication Number: CN 201234298 Y  
Publication Date: 2009-05-06  
Produced: 2011-04-17  
Updated: 2011-04-15

**Expanding apparatus for smart phone**

Abstract The utility model relates to an extension device of a smart cellphone, which comprises an interface

一种智能手机的扩展装置 Abstract  
本实用新型涉及一种智能手机的扩展装置,包括用于与智能手机进行数据交互的接口模块、与所述接口模块相连的数据接口处理单元、以及I/O接口单元,所述接口模块处理单元还通过I/O接口数据引线与I/O接口单元相连,所述I/O接口单元连接有输入输出设备或接口。本实用新型采用智能手机与扩展装置的分离设计,通过将智能手机作为扩展装置的主控模块,让智能

タイトルなど、見つけやすい部分で、英語で検索し、対応する人間翻訳の中国語キーワードをピックアップ  
(中国語が分からなくても選択可能)

CPA  
Global Discover  
国内代理店  
日本技術貿易(株)

図 6 CPA Global Discoverでの検索結果表示例

## 2.7 出願人 中国語表記取得方法例 CNIPR 中国語DBより

CNIPR中国語DBより（無料ゲストID可）

<http://search.cnipr.com/>

出願人（申請人）欄の虫メガネ部をクリックし、中国語会社名をコピー＆ペーストして入力し、異表記を取り出すこともできる。

The screenshot shows the CNIPR search platform interface. At the top, there are search filters for patent types (e.g., 中国发明专利, 中国实用新型) and countries (e.g., 美国, 日本, 英国). Below these are search options like 二次检索 and 过滤检索. The main search area has a text input field labeled '请输入申请人' with '富士胶片' entered, and a '检索' button. To the right, there are search criteria fields (A: 申请(专利)号, B: 申请人, C: 公开(公告)号, D: 公开, E: 名称, F: 摘要, G: 主分类号, H: 分类号, I: 申请(专利权)人, J: 发明人) with search icons. The search results are divided into two sections: '公司代码检索结果' and '申请人检索结果'. The '公司代码检索结果' section shows a list of companies related to '富士胶片', with '富士胶片' and '富士胶片; 亨特' highlighted. The '申请人检索结果' section shows a list of companies including '富士胶片亨特化学制品美国有限公司', '富士胶片欧洲制造公司', '富士胶片迪麦提克斯公司', and '苏州富士胶片映像机器有限公司'.

図7 CNIPR中国語DBからの中国出願人名 異表記取得方法例

# 3.中国特許収録確認

## 3.1収録確認方法

- PatBase、Orbit.com、Discoverについて、それぞれのDBの機能異なり、同じ条件で評価できず、以下の方法で収録を確認

表 2 DB収録確認方法 (2011.10.15確認)

DB	英語要約収録	中国語要約収録
PatBase	コマンド(AB=YES)	目視確認(各n=20)
Orbit.com	コマンド (AB=YES) or (MTAB=YES)	コマンド (OAB=YES)
Discover	汎用語検索	目視確認(各n=20)

英語(含む機械翻訳)要約があるもののみ、有効収録と考え、収録確認実施。

(中国語要約収録確認は、英語要約あるものについて収録率算出)

Discover 汎用語検索式

AB:("com\*" or "con\*" or "pro\*" or "inv\*" or "met\*")

(注)汎用語を用いた検索 (参照)

2009年11月情報プロフェショナルシンポジウム  
“中国・台湾および韓国特許庁データベースの全文検索機能とその活用” (伊藤氏他)<sup>12</sup>

## 3.2 中国公開特許 1990-2010年発行分 英語・中国語要約収録確認

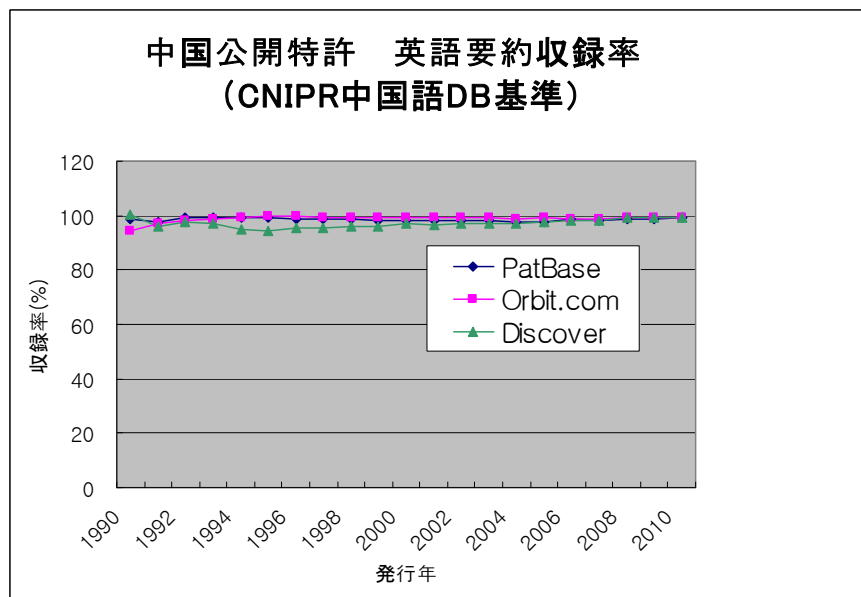


図 8 中国公開特許 英語要約収録

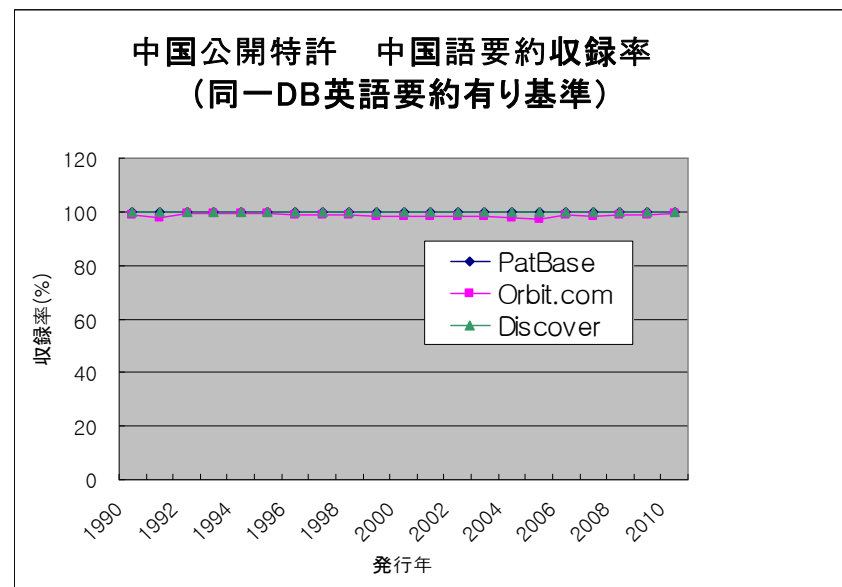


図 9 中国公開特許 中国語要約収録

2011.10.15確認

3DBとも、1990-2010年発行公開特許は、英語要約及び中国語要約とも、実用上問題ないレベルで収録されている。(CNIPR中国語DBとの比較)



# 3.3 中国公開特許 2011年発行分 収録タイムラグ確認(英語、中国語要約)

2011.10.15確認

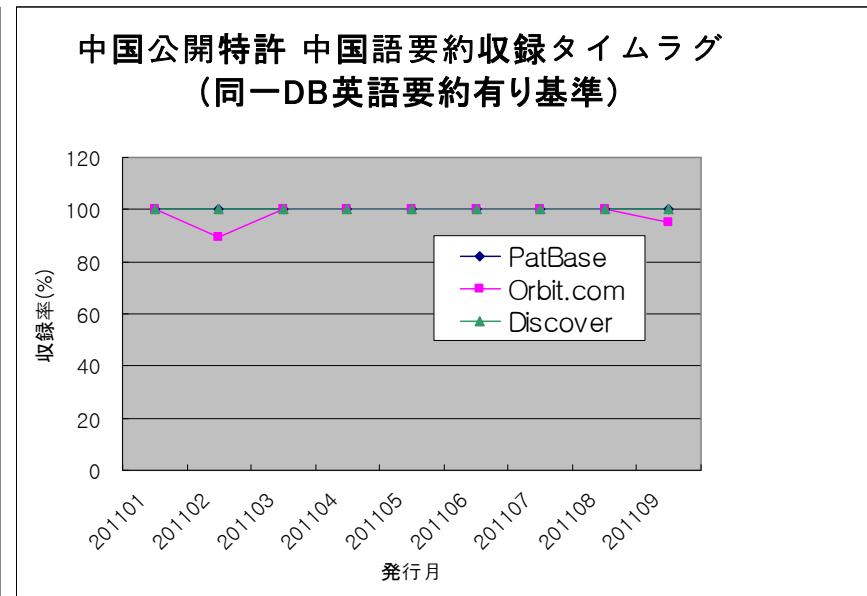
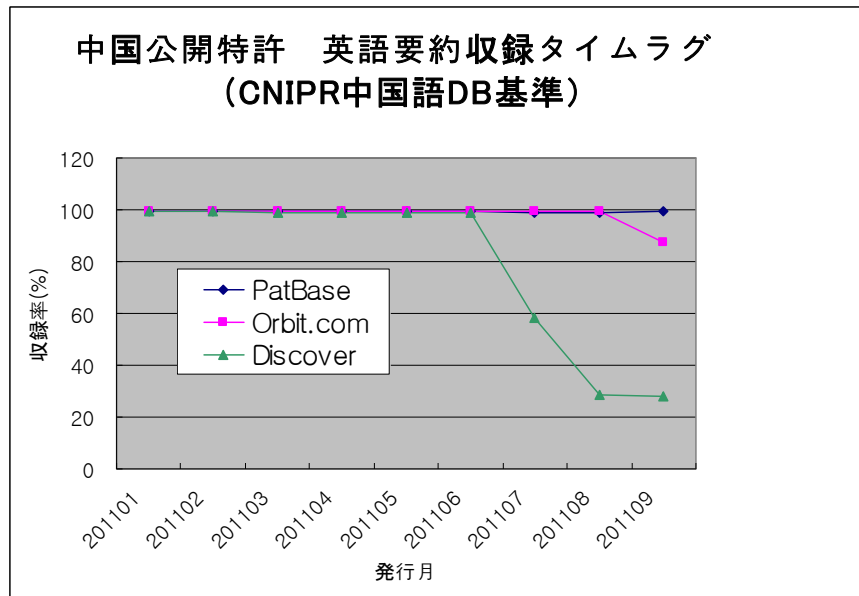


図 10 中国公開特許 英語要約タイムラグ

図 11 中国公開特許 中国語要約タイムラグ

PatBase,Orbit.comは中国公開特許について英語、中国語とも収録タイムラグほとんどない。Discoverは英語要約完全収録まで約3ヶ月のタイムラグ有り。(ただし、Discoverも中国語分は公開後2週間程度で完全収録確認済み)

また、英語要約の収録あれば、中国語要約もほぼ収録されている。

# 3.4 中国实用新案 1990-2010年発行分 英語・中国語要約収録確認

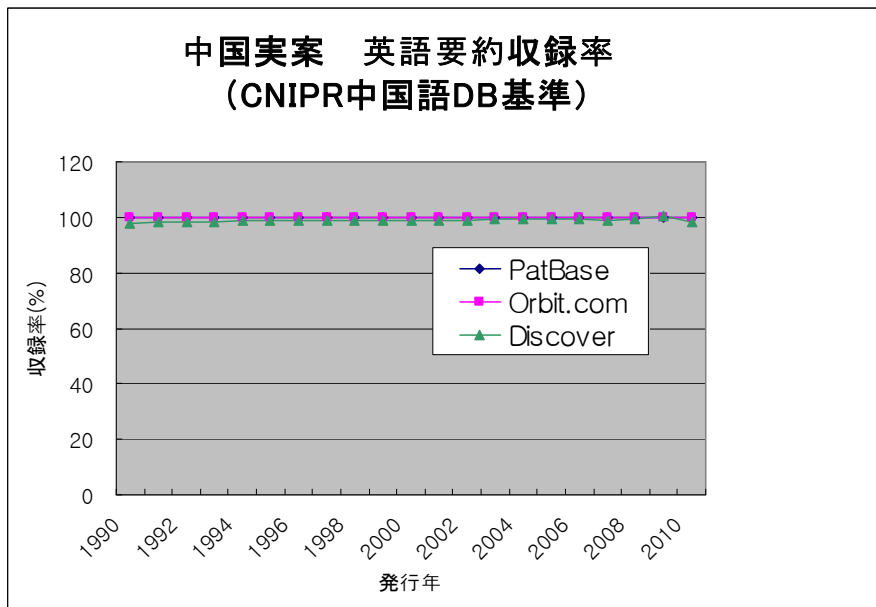


図 12 中国实用新案 英語要約収録

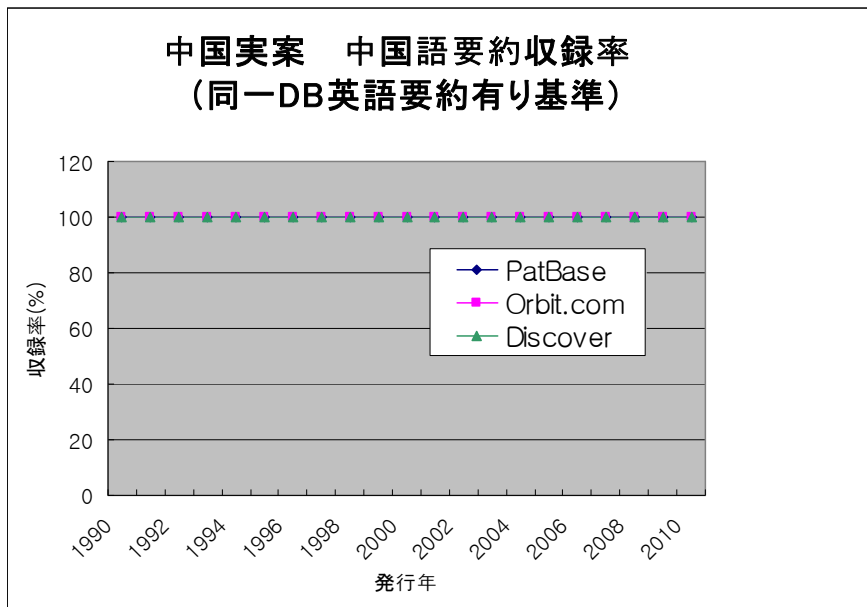


図 13 中国实用新案 中国語要約収録

2011.10.15確認

3DBとも、1990-2010年発行实用新案は、英語要約及び中国語要約とも、実用上問題ないレベルで収録されている。(CNIPR中国語DBとの比較)

# 3.5 中国実用新案 2011年発行分 収録タイムラグ確認(英語、中国語要約)

2011.10.15確認

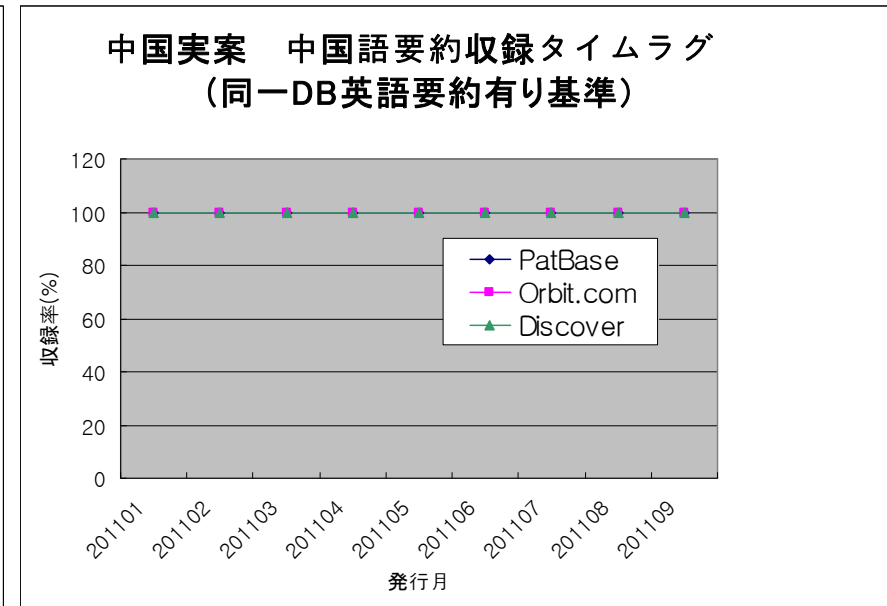
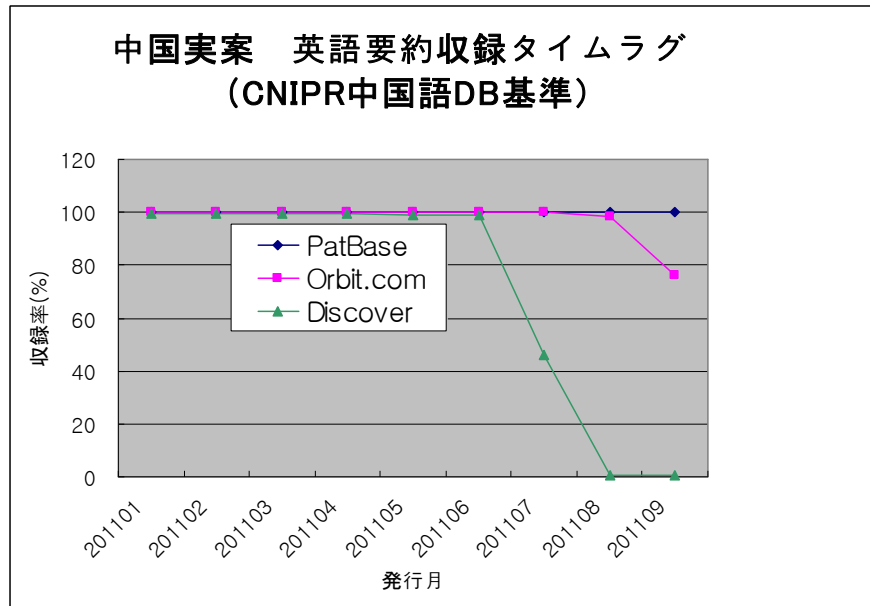


図 14 中国実用新案 英語要約タイムラグ 図 15 中国実用新案 中国語要約タイムラグ

公開特許同様で、PatBase,Orbit.comは中国実用新案について英語、中国語とも収録タイムラグほとんどない。Discoverは英語要約完全収録まで約3ヶ月のタイムラグ有り。(ただし、Discoverも中国語分は公開後2週間程度で完全収録確認済み)

また、英語要約の収録あれば、中国語要約もほぼ収録されている。



## 3.6 中国特許、実用新案収録まとめ

- PatBase, Orbit.com, Discoverとも、中国公開特許、実案についての収録は、実用上問題ないレベルまで収録されている。ただし、Discoverについては、英語要約完全収録まで約3ヶ月のタイムラグ有り。ただし、3DBとも中国語についてはほとんどタイムラグなく収録されている
  - 英語要約(含む機械翻訳)が収録されていれば、中国語要約もほぼ収録されている。
    - 英語&中国語の検索可能な収録はされている！
    - (目視で見える限り、中国語要約あれば、中国語クレーム以下もほぼ収録されていた)
- (注)3DBとも、ファミリータイプのDBなので、中国以外のデータ収録(英語要約有り等)でカバーされた可能性もあり。

# 4.出願人検索検証

## 4.1 中国出願人の検索検証

- 中国人出願人の検索の場合、同じ出願人でも、複数の英語表記ある場合多く、網羅的に検索するのは難しい。

例として、“乐凯集团第二胶片厂”

(英語表記例: **SECOND FILM FACTORY OF LUCKY**)

を取り上げ、3DBにて、“英語のみ”、“英語+中国語”での出願人検索を実施し、ハイブリッド検索の効果を検証。

英語検索式例: **(SECOND or 2 or 2.) and (FILM) and (LUCKY)**

中国語検索式例: 乐凯集团第二

注: CPA Global Discoverの場合は、中国語検索時もストリングサーチ(部分一致)ではヒットせず、ワイルドカードを利用して検索

## 4.2 中国出願人のハイブリッド検索効果

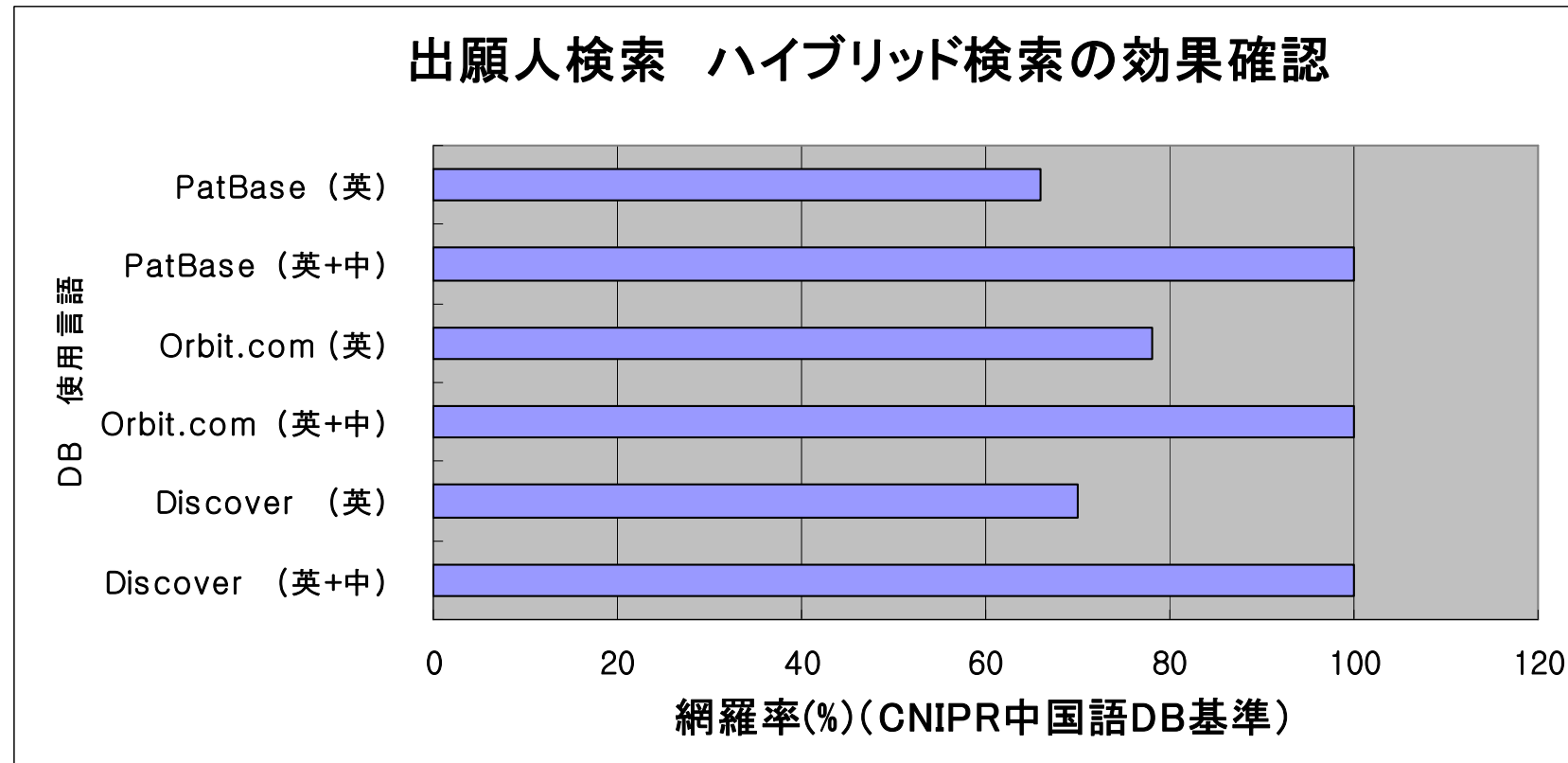


図 16 中国出願人検索 ハイブリッド検索の効果確認

いずれのDBでも英語出願人名での検索ではもれが発生し、中国語出願人名での検索併用で、CNIPR中国語DB同等に網羅できた。

2011.10.9確認

## 4.3 英語出願人名の解析例

表 3 英語出願人名表記割れ解析例 (50出願中)

英語出願人名	件数
SECOND FILM FACTORY OF LUCKY G	26
SECOND FILM FACTORY OF LUCKY GROUP	7
NO 2 FILM FACTORY LEKAI GROUP	4
NO 2 FILM FACTORY LUCKY GROUP	3
LEKAI GROUP NO 2 FILM FACTORY	2
NO 2 FILM FACTORY OF LEKAI GRO 2	2
NO 2 FILM FACTORY LUCKY GORUP	1
英語出願人情報なし	10

2011.10.9 確認

LUCKYの表記が社名の中国語読み(ピンイン)からくる“LEKAI”になっている例があり。

また、そもそも英語出願人情報も、全体の2割にあたる10件欠落あり、英語による出願人検索ではヒットできない。



中国特許 出願人検索では、中国語併用による検索が必須と考える。

# 5. キーワード検索検証

## 5.1 キーワード検索時の注意点

- 英語←→ 中国語 の訳(含む機械翻訳)の違い。  
(意訳、誤訳によるものなど)
- データ収録によるもの  
(例: 英語情報、中国語情報の欠落等)

本検討では、日本語での“**干渉縞**”を例に取り上げ、英語及び中国語でのキーワード検索について評価。

## 5.2 キーワード選定

表 4 日本語“干渉縞”の翻訳例

### 英語訳例

fringe pattern  
interference fringe  
interference band  
interference pattern  
interference stripe  
Moire fringe

### 中国語訳例

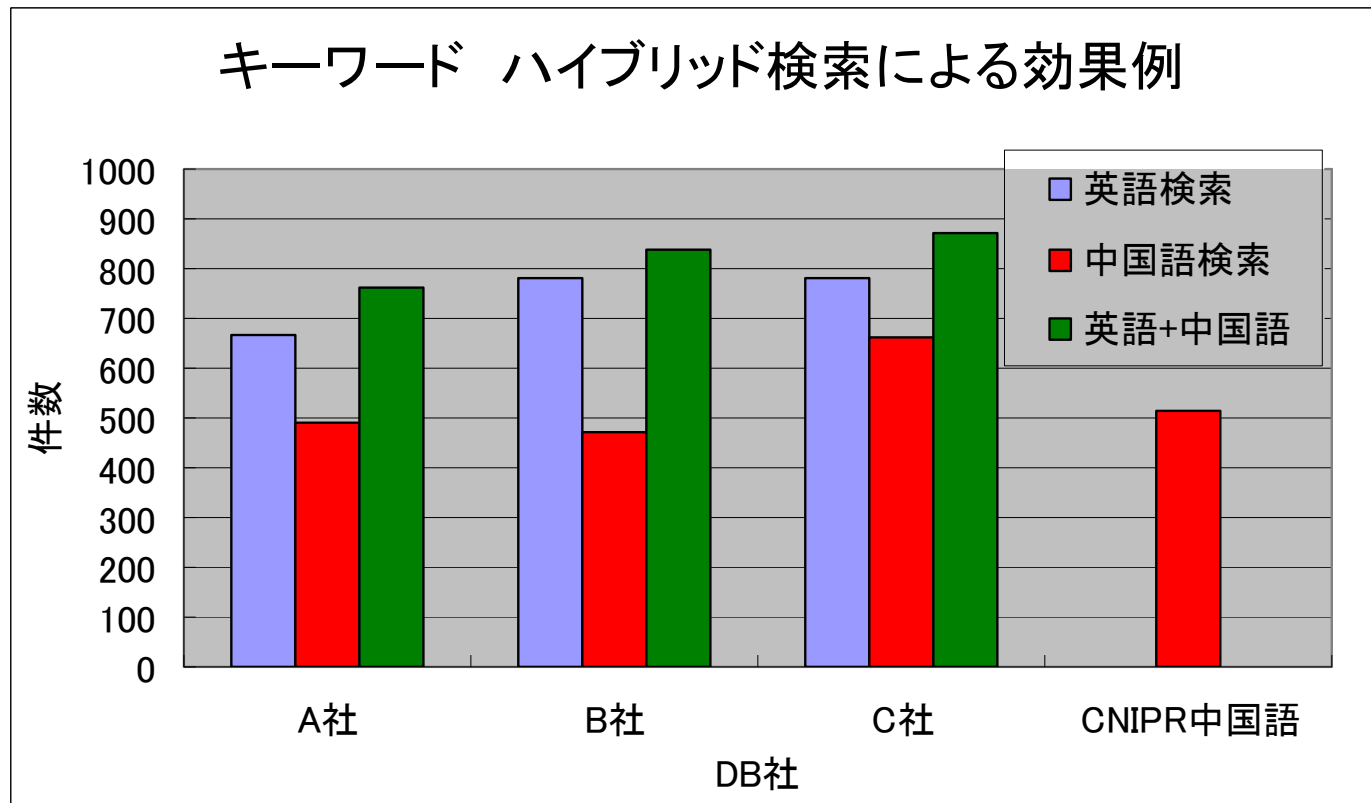
干涉条纹  
干涉频带  
干涉光带  
干涉带  
干扰条纹  
干扰频带  
干扰光带  
干扰带

(注) 英語については複数系等を考慮し、  
末尾をワイルドカード処理して検索。  
Discoverについては中国語も  
末尾ワイルドカード処理を実施

→これらの英語、中国語キーワードを用いて、検索を実施

## 5.3 キーワードハイブリッド検索効果

1991年以降出願の中国公開特許について、タイトル・要約を対象に“干渉縞”を英語・中国語ハイブリッド検索



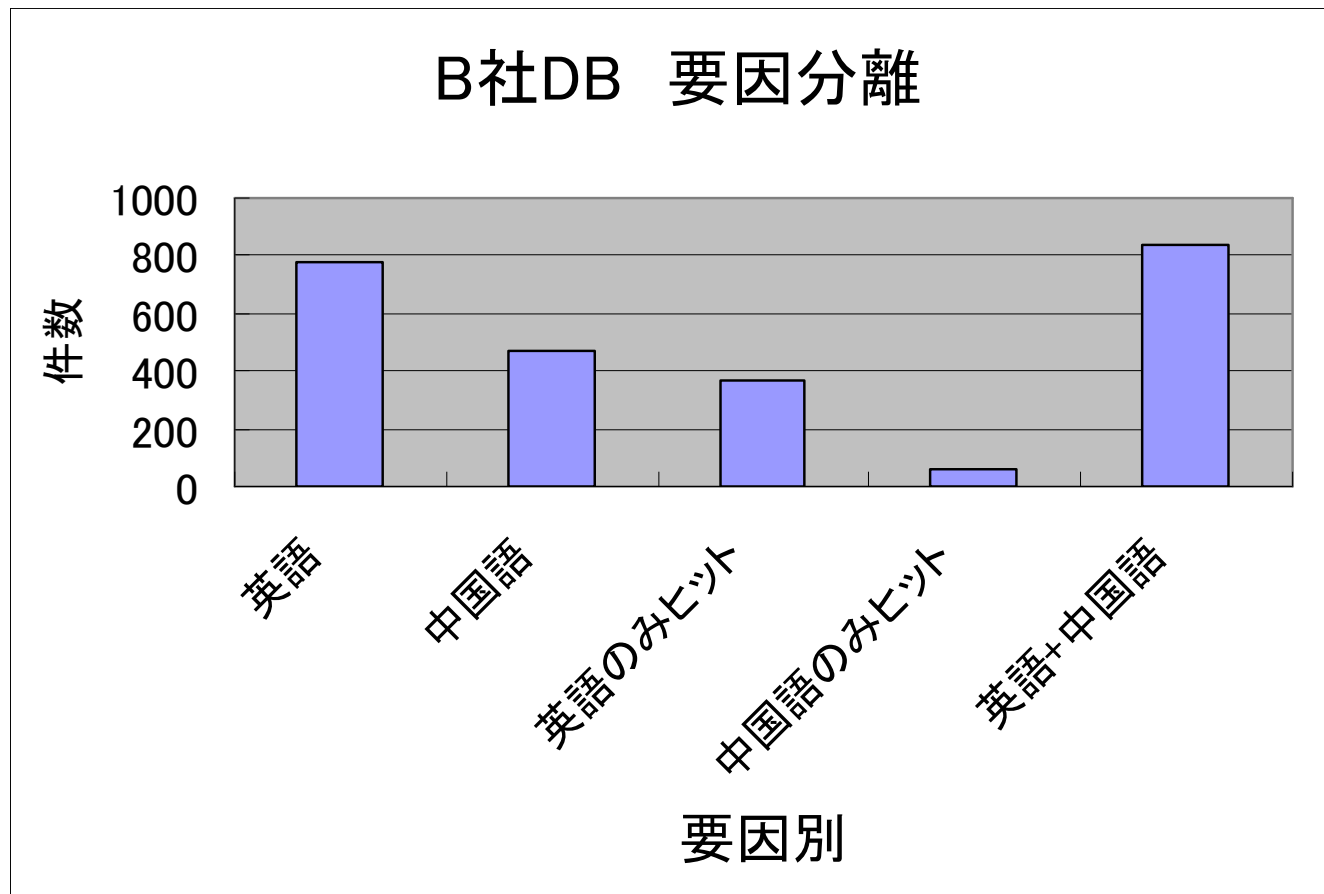
英語・中国語のハイブリッド検索により、単独での検索より、ヒット数が増大  
(網羅性↑)

2011.10.19確認

図 17 キーワード ハイブリッド 検索の効果確認

# 5.4 キーワード検索 要因分離

B社DBで、それぞれの言語検索での要因分離を実施



中国語は確実に該当するもののみチョイスしたが、英語、中国語のみでヒットするものもある。

2011.10.19確認

図 18 英語検索、中国語検索のヒット要因分離



## 5.5 英語検索で漏れた例解析

表 5 英語検索でのみれ解析

halftone image 英語記載なし	“干涉条纹”が意識で変化 “干涉条纹”が意識で消える
plot fringe	“干扰带”が意識で変化
interference strips	“干涉条纹”のstripeが、strips に訳されている。
interferometric fringes	“干涉条纹”の訳があまり使われ ない形で英訳されている。
interference ripple	“干涉条纹”の模様部がrippleに訳され ている。

誤訳というよりは、言語体系の違いによる意識の部分で、単語だけでは想定できない訳の場合あり、注意が必要と考える。

## 5.6 中国語検索で漏れた例解析

表 6 中国語検索で漏れた例解析

莫尔条纹	(モアレ模様)
干扰特征	(干涉特徴)
干涉特性	(干涉特性)
干扰图	(干涉図)
干涉图样	(干涉模様)
干涉图案	(干涉模様)
中国語収録なし	(データ欠落)

中国語は異表記多く、すべてをカバーするのは難しい。異表記を随時キーワード集積していくことが重要と考える。

## 6. まとめ

- 英語・中国語ハイブリッド検索可PatBase,Orbit.com(QPAT),Discoverについて、中国公開特許、実用新案の収録を確認し、英語だけでなく中国語データ(要約まで確認済み)が収録されていることを確認。

(実用上問題ないレベル)

- 英語、中国語を併用して検索することにより網羅性上がり、CNIPR中国語DBを用いて補完するレベルと同等の効果あり、かつ一つのDB内で処理できることから、その工数対効果は高いと考える。
- 中国語出願人検索に於いては、英語だけでなく、中国語も併用することによる効果が特に大きい。

このようなハイブリッド検索可能な機能が、いろいろなDBベンダーから提供予定と聞いており、今後のトレンドの一つと考える。

ご清聴ありがとうございました。

谢谢，再见！

Thank you !

