

B12

適合率を重視した特許調査支援

アジア特許情報研究会

○花王株式会社

はやぶさ国際特許事務所

安藤俊幸

桐山 勉

発表内容

テキストマイニングによるKW分析を活用した特許調査支援

再現率(網羅性)重視で検索集合作成(前提) → ノイズ増加
適合率(探したいもの)重視で抽出/ソート → スクリーニング支援ツール/方法

○KW分析(対象言語:中国語・日本語)

○中国語(その他言語)のネットワーク分析とその応用

検討内容

- ・多言語対応のネットワーク分析/可視化ツールとしてCytoscapeを検討
- ・調査に重要(特徴的)なKWを網羅的に抽出する → 専門用語辞書(INDEX)
- ・中国語/日本語解析ツール → PatAnalyzer(新名称)
- ・文書クラスタリングへの応用 → 類似率(他項目)による公報のネットワーク
- ・類似文書抽出のための類似度、重み付け方法 → 類似検索(適合率重視)

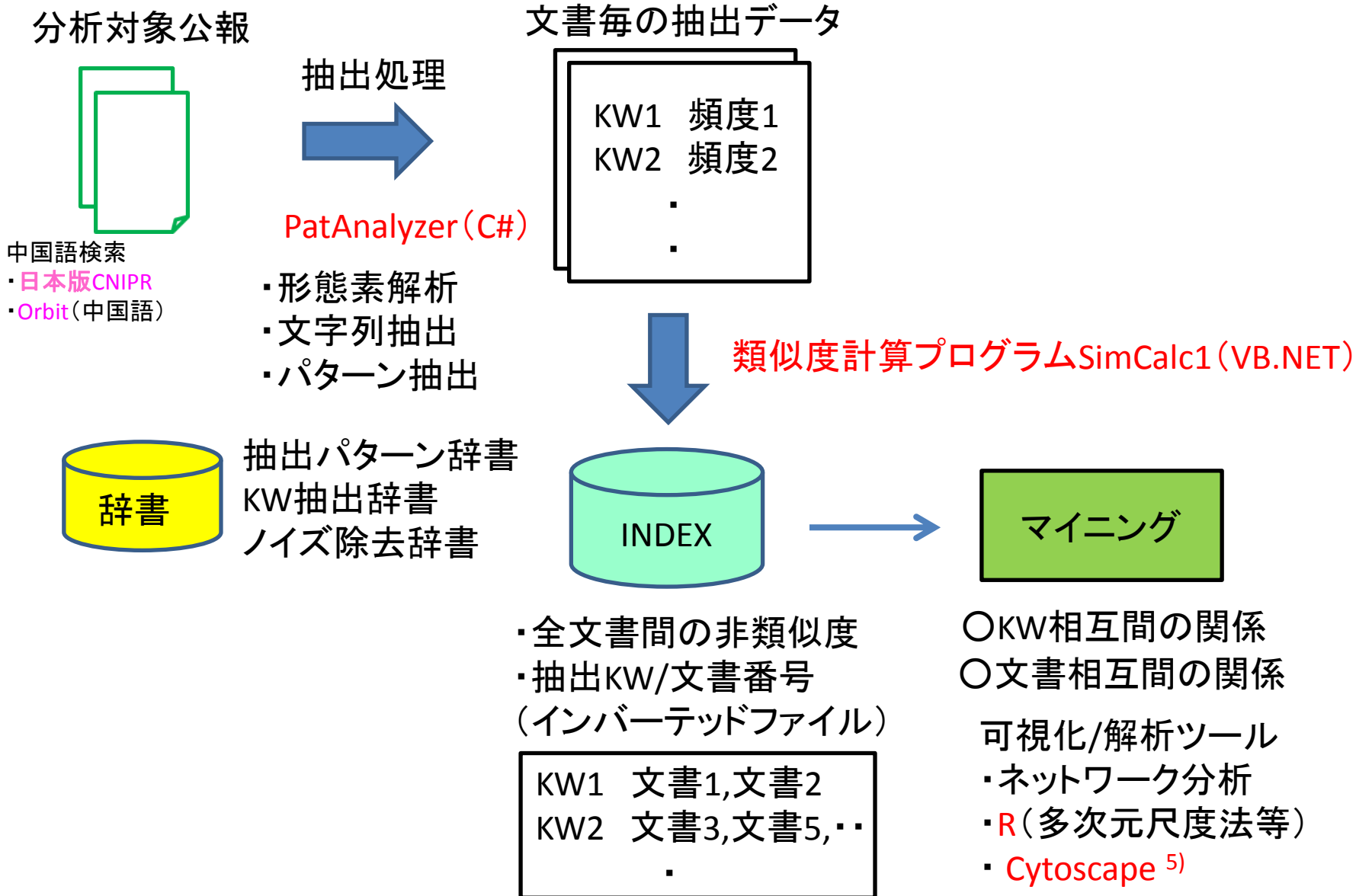
検討対象データベース

- ・日本版CNIPR
- ・Orbit.com

検討対象分野

- ・iPS細胞

テキストマイニングの処理の流れ(概要)



技術分野 本発明は、分化した体細胞を初期化して**誘導多能性幹細胞**を誘導する作用を有する**核初期化因子**に関するものである。

解析対象

表題 核初期化因子

要約 胚やES細胞を利用せずに分化細胞の初期化を誘導し、ES細胞と同様な**多能性**や**増殖能**を有する**誘導多能性幹細胞**を簡便かつ再現性よく樹立するための手段として、下記の3種の遺伝子：**Octファミリー遺伝子**、**Klfファミリー遺伝子**、及び**Mycファミリー遺伝子**の各遺伝子産物を含む**体細胞の核初期化因子**が提供される。

請求項1 体細胞の**核初期化因子**であって、下記の3種類の遺伝子：**Octファミリー遺伝子**、**Klfファミリー遺伝子**、及び**Mycファミリー遺伝子**の各遺伝子産物を含む因子。

請求項2 下記の3種の遺伝子：Oct3/4、Klf4、及びc-Mycの各遺伝子産物を含む請求項1に記載の因子。

請求項3 下記の遺伝子：**Soxファミリー遺伝子**の遺伝子産物をさらに含む請求項1又は2に記載の因子。

請求項4 Sox2の遺伝子産物を含む請求項3に記載の因子。

請求項5 Mycファミリー遺伝子の遺伝子産物とともに、あるいはMycファミリー遺伝子の遺伝子産物に換えてサイトカインを含む請求項1ないし4のいずれか1項に記載の因子。

請求項6 サイトカインがbFGF及び/又はSCFである請求項5に記載の因子。

請求項7 下記の遺伝子：TERT遺伝子の遺伝子産物をさらに含む請求項1ないし6のいずれか1項に記載の因子。

請求項8 下記の遺伝子：SV40 Large T antigen、HPV16 E6、HPV16 E7、及びBmi1からなる群から選ばれる1種以上の遺伝子の遺伝子産物をさらに含む請求項1ないし7のいずれか1項に記載の因子。

請求項9 Fbx15、Nanog、ERas、ECAT15-2、Tcl1、及びβ-cateninからなる群から選ばれる1種以上の遺伝子の遺伝子産物をさらに含む請求項1ないし8のいずれか1項に記載の因子。

請求項10 ECAT1、Esg1、Dnmt3L、ECAT8、Gdf3、Sox15、ECAT15-1、Fth17、Sall4、Rex1、UTF1、Stella、Stat3、及びGrb2からなる群から選ばれる1種以上の遺伝子の遺伝子産物をさらに含む請求項1ないし9のいずれか1項に記載の因子。

請求項11 体細胞の核初期化により誘導多能性幹細胞を製造する方法であって、体細胞に対して請求項1ないし10のいずれか1項に記載の核初期化因子を接触させる工程を含む方法。

請求項12 体細胞がヒト細胞である請求項11に記載の方法。

請求項13 請求項11又は12に記載の方法により得られた誘導多能性幹細胞。

請求項14 請求項13に記載の誘導多能性幹細胞から分化誘導された体細胞。

請求項15 細胞の分化能及び/又は増殖能を改善する方法であって、細胞に対して請求項1ないし10のいずれか1項に記載の核初期化因子を接触させる工程を含む方法。

請求項16 細胞がヒト細胞である請求項15に記載の方法。

請求項17 請求項15又は16に記載の方法により得られた細胞。

請求項18 請求項17に記載の細胞から分化誘導された体細胞。

解析対象

核重新编程因子

作为用于不利用胚和ES细胞而诱导分化细胞的重新编程，简便且再现性强地建立具有与ES细胞同样的**多能性**和**增殖能力**的**诱导式多能性干细胞**的方法，本发明提供了含有下述3种基因：**Oct家族基因**、**Klf家族基因**和**Myc家族基因**的各基因的产物的体细胞的核重新编程因子。

1. 一种体细胞的**核重新编程因子**，该因子含有下述3种基因：**Oct|**家族基因、**K1f**家族基因和**Myc**家族基因的各基因的产物。
2. 权利要求1中记载的因子，其含有下述3种基因：**Oct3/4**、**K1f4|**和**c-Myc**的各基因的产物。
3. 权利要求1或2中记载的因子，其还含有下述基因：**Sox**家族**|**基因的基因产物。
4. 权利要求3中记载的因子，其含有**Sox2**的基因产物。
5. 权利要求1至4中任意一项中记载的因子，其含有细胞因子，**|**该细胞因子和**Myc**家族基因的基因产物共同存在、或者代替**Myc**家族**|**基因的基因产物。
6. 权利要求5中记载的因子，所述细胞因子为**bFGF**和/或**SCF**。
7. 权利要求1至6任意一项中记载的因子，其还含有下述的基因：**|TERT**基因的基因产物。
8. 权利要求1至7任意一项中记载的因子，其还含有选自于由下**|**述的基因构成的组中的1种以上的基因的基因产物：**SV40大T抗原**、**|HPV16 E6**、**HPV16 E7**和**Bmi1**。
9. 权利要求1至8任意一项中记载的因子，其还含有选自于由下**|**述的基因构成的组中的1种以上的基因的基因产物：**Fbx15**、**Nanog**、**|ERas**、**ECAT15-2**、**Tc11**和 **β -连环素**。
10. 权利要求1至9任意一项中记载的因子，其还含有选自于由**|**下述的基因构成的组中的1种以上的基因的基因产物：**ECAT1**、**Esg1**、**|Dnmt3L**、**ECAT8**、**Gdf3**、**Sox15**、**ECAT15-1**、**Fth117**、**Sa114**、**Rex1**、**|UTF1**、**Ste11a**、**Stat3**、和**Grb2**。
11. 通过体细胞的核重新编程制备诱导式多能性干细胞的方法，**|**其包括使权利要求1至10任意一项中记载的核重新编程因子与该体细**|**胞接触的步骤。
12. 权利要求11中记载的方法，所述体细胞为人细胞。
13. 由权利要求11或12中记载的方法获得的诱导式多能性干细**|**胞。
14. 通过诱导权利要求13中记载的诱导式多能性干细胞分化而得**|**到的体细胞。
15. 一种改善细胞的分化能力和/或增殖能力的方法，该方法包括**|**使权利要求1至10任意一项中记载的核重新编程因子与细胞接触的步**|**骤。
16. 权利要求15中记载的方法，所述细胞为人细胞。
17. 由权利要求15或16中记载的方法获得的细胞。
18. 通过诱导权利要求17中记载的细胞分化而得到的体细胞。

PatAnalyzer 中国語/日本語KW解析ツール

IKAnalyzerNet²⁾を基にMecab(和布蕪:形態素)、Cabocha(係り受け)解析機能を追加
(中国語分詞ツール) (saezuri lite⁴⁾を介して呼び出し

入力

正規表現による
サーチ、抽出
機能追加

PatAnalyzer Ver.1.10

核重新編程因子|基因

正規表現

文字列サーチ 戻す サーチ 抽出

文抽出 文末:改行

19文

文字色 背景色 色設定

コピー

分詞開始

隣接頻度

単語	隣接頻度
核重	8
編	8
程	7
因子	7
作为	5
用于	5
不利	5
胚	5
es	4
細胞	4
诱导	4
分化	4
細胞	3
重新	3

分析用(文単位)

分詞出力(類似率)

日本語解析用

和布蕪解析

形態素 専門用語 形態素+専門用語

和布蕪

出力(類似率)

Cabocha

統計出力

中国語解析用

文字数:1127 時間:545ms 分詞数:537 効率(詞/秒):985

分詞
出力

saezuri lite

PatAnalyzer 中国語/日本語KW解析ツール

IKAnalyzerNet²⁾を基にMecab(和布蕪:形態素)、Cabocha(係り受け)解析機能を追加
(中国語分詞ツール) (saezuri lite⁴⁾を介して呼び出し)

入力

正規表現による
サーチ、抽出
機能追加

PatAnalyzer Ver.1.10

核初期化因子
胚やES細胞を利用せずに分化細胞の初期化を誘導し、ES細胞と同様な多能性や増殖能を有する誘導多能性幹細胞を簡便かつ再現性よく樹立するための手段として、下記の3種の遺伝子:Octファミリー遺伝子、Klfファミリー遺伝子、及びMycファミリー遺伝子の各遺伝子産物を含む体細胞の核初期化因子が提供される。

請求項1 体細胞の核初期化因子であって、下記の3種類の遺伝子:Octファミリー遺伝子、Klfファミリー遺伝子、及びMycファミリー遺伝子の各遺伝子産物を含む因子。

請求項2 下記の3種の遺伝子:Oct3/4、Klf4、及びc-Mycの各遺伝子産物を含む請求項1に記載の因子。

請求項3 下記の遺伝子:Soxファミリー遺伝子の遺伝子産物をさらに含む請求項1又は2に記載の因子。

正規表現

文字列サーチ 戻す サーチ 抽出

文抽出 文末:改行

925 語 358ms

文字色 背景色 色設定

コピー

隣接頻度

日本語解析用 saezuri lite

和布蕪解析

形態素

専門用語

形態素+専門用語

和布蕪

出力(類似率)

Cabocha

統計出力

分析用(文単位)

分詞出力(類似率)

中国語解析用

分詞開始

核	名詞,一般****核,カク,カク	Mycファミリー遺伝子	遺伝子産物 4
初期	名詞,一般****初期,シヨキ,シヨキ	遺伝子	遺伝子産物 3
化	名詞,接尾,サ変接続****化,カカ	Octファミリー遺伝子	Klf
因子	名詞,一般****因子,インシ,インシ	ファミリー遺伝子	2
EOS		分化誘導	体細胞 2
胚	名詞,一般****胚,ハイ,ハイ	種	遺伝子 2
や	助詞,並立助詞****や,ヤ,ヤ	工程	方法 2
ES	名詞,一般****	遺伝子	Octファミリー遺伝子 2
細胞	名詞,一般****細胞,サイボウ,サイ	接触	工程 2
ポー		方法	細胞 2
を	助詞,格助詞,一般***を,ヲ		
利用	名詞,サ変接続****利用,リヨウ,リョウ		
サ	動詞,白六****サ,サ,フル,去来コ接続オ		

処理文数=19 KW抽出=75 処理時間: 374ms

解析
出力

Orbit.comのコンセプト表示(英語)wo2007069666

418 RESISANT COLONY | 418 RESISTANT COLONY | 418 RSISTANT COLONY | ABNORMAL PHENOTYPE | AUCTION | BASIC FIBROBLAST GROWTH
FACTOR | BCF | BETA CATENIN | BETA EMBRYO | **BFGF** | BLASTOCYSTS CELL MASS | BMIL | BORTVIN ETAL | C MYC GENE | CARDIAC
INSUFFICIENCY | CARDIAC MUSCLE | CAUSING REJECTION PROBLEM | CELL GROWTH ABILITY | CELL STRAIN | CELL TRANSPLANTATION |
CELLULAR AUTO TRANSPLANTATION | CHIMERIC MOUSE MATE | CLONED EMBRYO | COLONY FORMATION OBSERVATION | COMPOUND TOXICITY | CYTOKINE |
DERMAL FIBROBLAST | DIFFERENTIATED CELL MORPHOLOGY | DIFFERENTIATED CELL NUCLEUS | DIFFERENTIATED CELL REPROGRAMING | DIFFERENTIATION
ABILITY | DIFFERENTIATION MAINTENANCE | DIFFERENTIATION PLURIPOTENCY INDUCTION | DISEASE TREATMENT THERAPEUTIC | **ECAT** | EGFP-IRES-PURO
CASSETTE | EGG | EGG ABSENCE | EMBRYO GONAD | EMBRYO SPINAL CORD | EMBRYONIC CELL PLURIPOTENCY | EMBRYONIC FIBROBLAST |
EMBRYONIC STEM CELL | ENDOGENOUS OCT | ESTABLISHED IP PICKED COLONIESNUMBER | **ETHICAL DIFFICULTY** | EXPERIMENTAL
ARRANGEMENT | FBX | FBX IP | FBX LIVER | FBX PROMOTER | FIBROBLAST REPROGRAMING | FTHL | FTHLL | GDF | **GENE KIND** | GENE OCT
FAMILY GENE | GENE OCT KIND | GENE TRANSDUCTION | GENOM IC CONTAMINATION | GENOM IC DNA | GENOM IC PCR ANALYSIS | GENOZMIC DNA
CONTAMINATION | GEO KNOCKIN | GEO MEFS | GEO MOUSE | GERM CELL TUMOR | GFP POSITIVE COLONY | GRB | GROWTH ABILITY | HEMOCYTE
CELL | HEPATIC CELL | HOMOMUTANT MOUSE | HPV | HUMAN CELL | INCREASING IP | **INDUCED PLURI INDUCING
DIFFERENTIATION** | INDUCED PLURI POTENT STEAM CELL | **INDUCED PLURI POTENT STEM CELL** | INDUCED PLURI
SELECTION | INSULIN DEPENDENT DIABETES | INTRODUCING GENE | IP CELL ESTABLISHMENT | **IP CELL INDUCING
DIFFERENTIATION** | ISOLATED GENE | **ISOLATED KLF** | ISOLATED MYC | ISOLATED OCT |
ISOLATED SOX | JUVENILE DIABETES | KINASE INHIBITOR | **KLF** | KLF FAMILY GENE | L MYC | MARKER GENE EXPRESSION | MEF CELL |
MEFS NUCLEUS | METHYLATED CPGS | MOUSE CELL | MOUSE DERIVED MEFS | MYC FAMILY | **MYC FAMILY GENE** | **MYC GENE
PRODUCT** | N MYC GENE | NANOG GENE | NANOG IP TRANSPLANTATION | NANOG-EGFP-IRES-PURO MOUSE | NCBI ACCESSION | OCT
FAMILY GENE | ORGAN TRANSPLANTATION | PHYSIOLOGICAL FUNCTION | PLURI MAINTENANCE | PLURIPOTENCY MAINTENANCE |
PRIMORDIAL GERM CELL | PROPERTY TAKING ADVANTAGE | PSEUDO PREGNANT MOUSE | PSEUDO UTERI | PUROMYCIN SELECTION | REJECTION
CAUSE | REPROGRAMED COLONY | REPROGRAMING EFFICIENCY VIEWPOINT | **REPROGRAMING FACTOR** | REPROGRAMING
INDUCTION ABILITY | REPROGRAMING INDUCTION LOSS | REPROGRAMING PHENOMENON EASY OBSERVATION | REPROGRAMING PHENOMENON ENABLE
OBSERVATION | RETRO VIRUS RECEPTOR | REX | SCF | **SOMATIC CELL NUCLEAR REPROGRAMING** | SOMATIC CELL NUCLEUS | SOMATIC CELL
REPROGRAMING | SOX | SOX FAMILY GENE | **SOX GENE PRODUCT** | SPINAL CORD INJURY | STAID CELL NUCLEAR REPROGRAMING | STELLA |
STEM CELL FACTOR | STO CELL | STRAIN CELL EXTRACT | TCL | TERT GENE GENE PRODUCT | TRANSPLANTATION THERAPY |
UNDIFFERENTIATED CELL MARKER | UNDIFFERENTIATED STATUS MAINTENANCE | UNDIFFERENTIATION MARKER | UTF |

英語のコンセプトだけでなく日本語、中国語表示できないか？

ワードクラウド検討 WO2007069666請求項

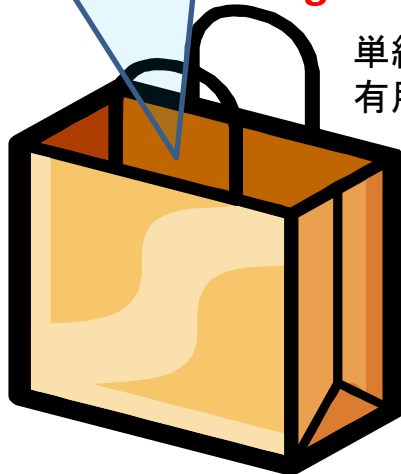
RMeCab⁷⁾ + wordcloud⁸⁾による
形態素レベルのwordcloud 名詞を抽出

PatAnalyzer + wordcloud による
専門用語レベルのwordcloud

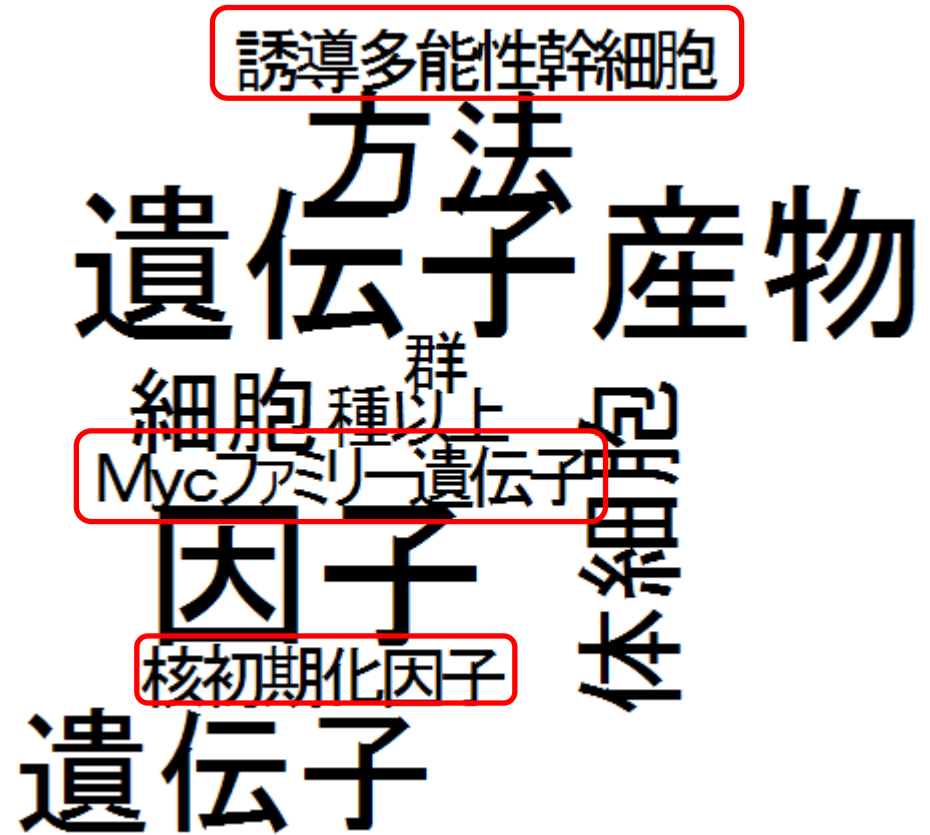


Bag-of-words モデル

単純なモデルの割には
有用だが限界も...



単語の出現頻度で
フォントサイズを規定



名詞の隣接頻度情報より
専門用語を抽出
適合性向上に有用

ワードクラウド検討 WO2007069666明細書全文

注)
RMeCab + wordcloud による
形態素レベルのwordcloud

PatAnalyzer + wordcloud による
専門用語レベルのwordcloud



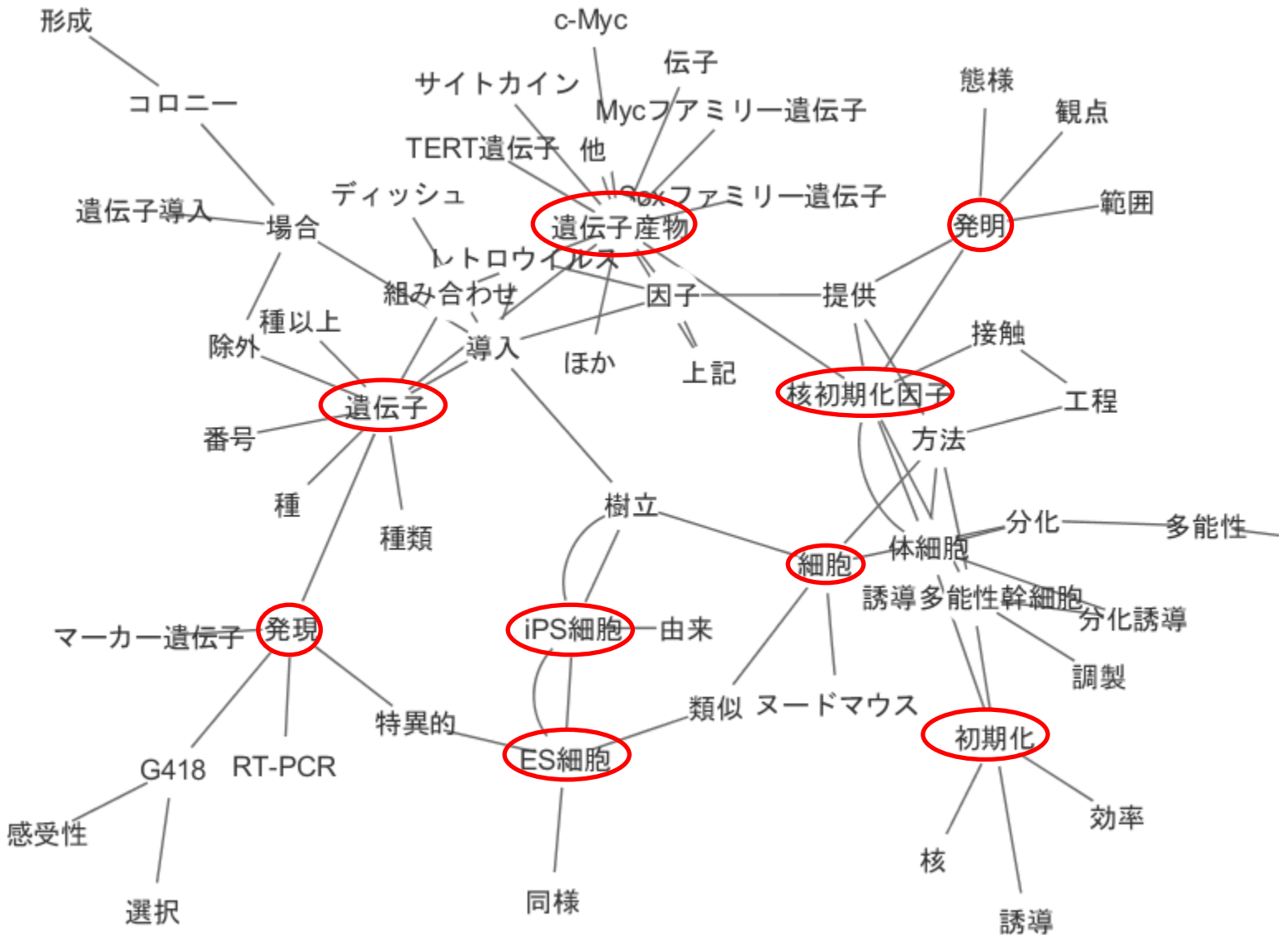
注) wordcloud : RのPackages
中国語の使用には中国語版Windows上に
Rをインストールする必要あり→難有り

薄まって抽出されない例: Mycファミリー遺伝子

Cytoscape⁵⁾による専門用語のネットワーク表示

WO2007069666明細書上位100語(専門用語)

Cytoscape3.1.0



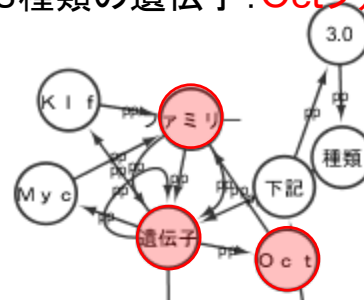
ネットワーク分析の観点からみた重要ノードの抽出→例: 媒介中心性 (Betweenness centrality)

Mecabによる形態素レベルの有向グラフ

請求項1 体細胞の核初期化因子であって、下記の3種類の遺伝子: Octファミリー遺伝子、Klfファミリー遺伝子、及びMycファミリー遺伝子の各遺伝子産物を含む

請求項1 体細胞の核初期化因子であって、下記の3種類の遺伝子: Octファミリー遺伝子、Klfファミリー遺伝子、及びMycファミリー遺伝子の各遺伝子産物を含む

。EOS



抽出例2 Oct-ファミリー-遺伝子
単語 n-gram n=3

抽出例1 核-初期-化-因子
単語 n-gram n=4

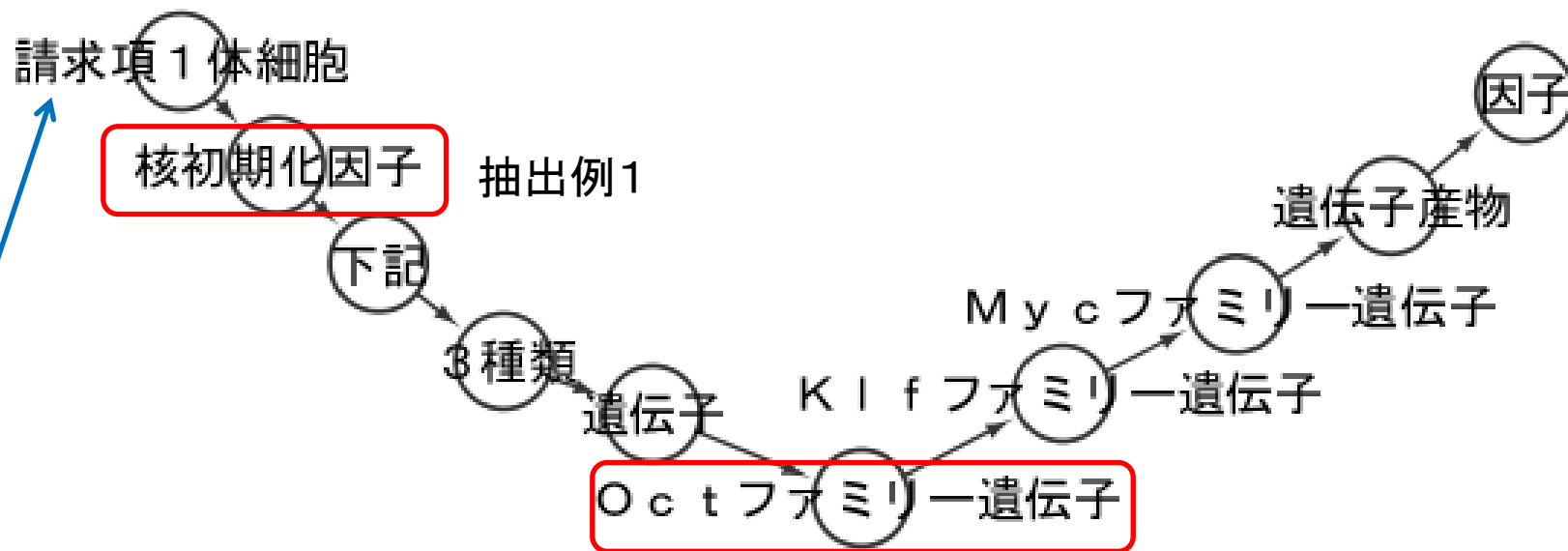


文頭より順に名詞を抽出

スタート

PatAnalyzerによる**専門用語**レベルの有向グラフ

請求項1 体細胞の核初期化因子であって、下記の3種類の遺伝子：
Octファミリー遺伝子、Klfファミリー遺伝子、及びMycファミリー遺伝子の
各遺伝子産物を含む因子



課題 特許用語の排除辞書
例: 請求項1

抽出例2

Cabochaによる係り受け解析と節関係解析

請求項1 体細胞の核初期化因子であって、下記の3種類の遺伝子: Octファミリー遺伝子、Klfファミリー遺伝子、及びMycファミリー遺伝子の各遺伝子産物を含む因子

Cabochaによる係り受け解析



Cabochaによる節関係解析



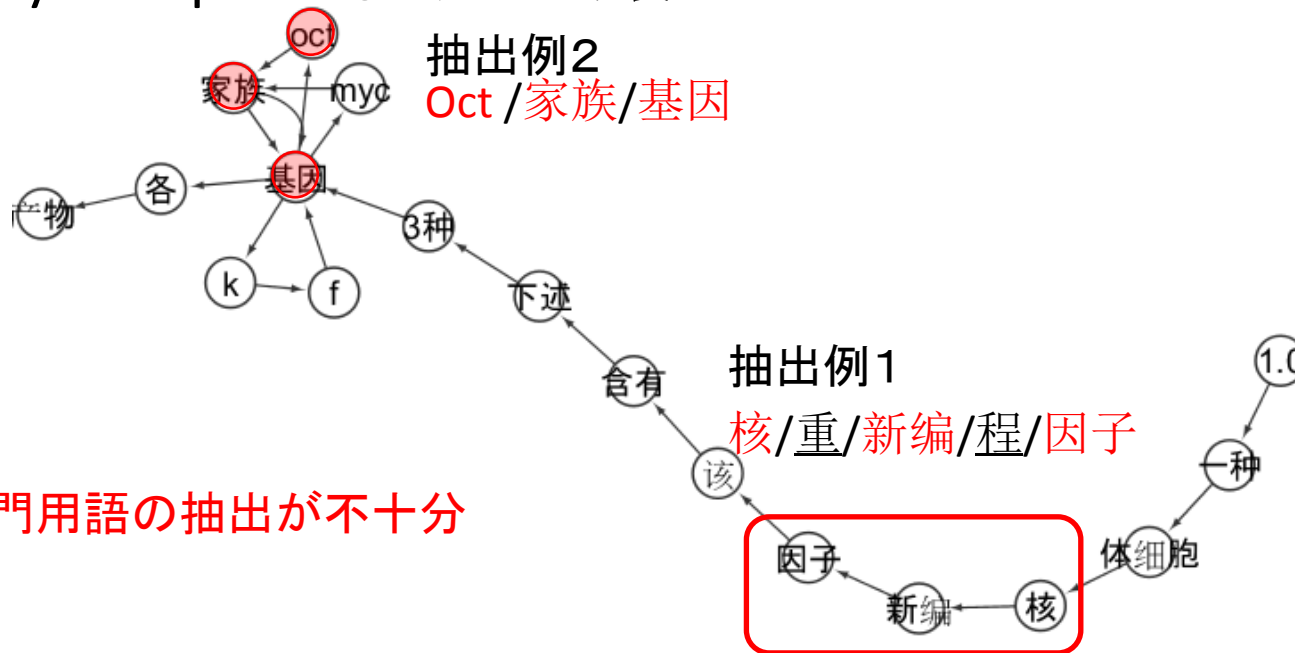
IKAnalyzer²⁾による分詞とネットワーク表示

CN101356270 請求項1

1. /一种/体细胞/的/核/重/新编/程/因子/, /该/因子/含有/
下述/3种/基因/: Oct/家族/基因、/K/1/f家族/基因/和/
Myc/家族/基因/的/各/基因的/产物/。

IKAnalyzerによる分詞 (/が分詞位置、下線は抽出されない)

Cytoscapeによるネットワーク表示



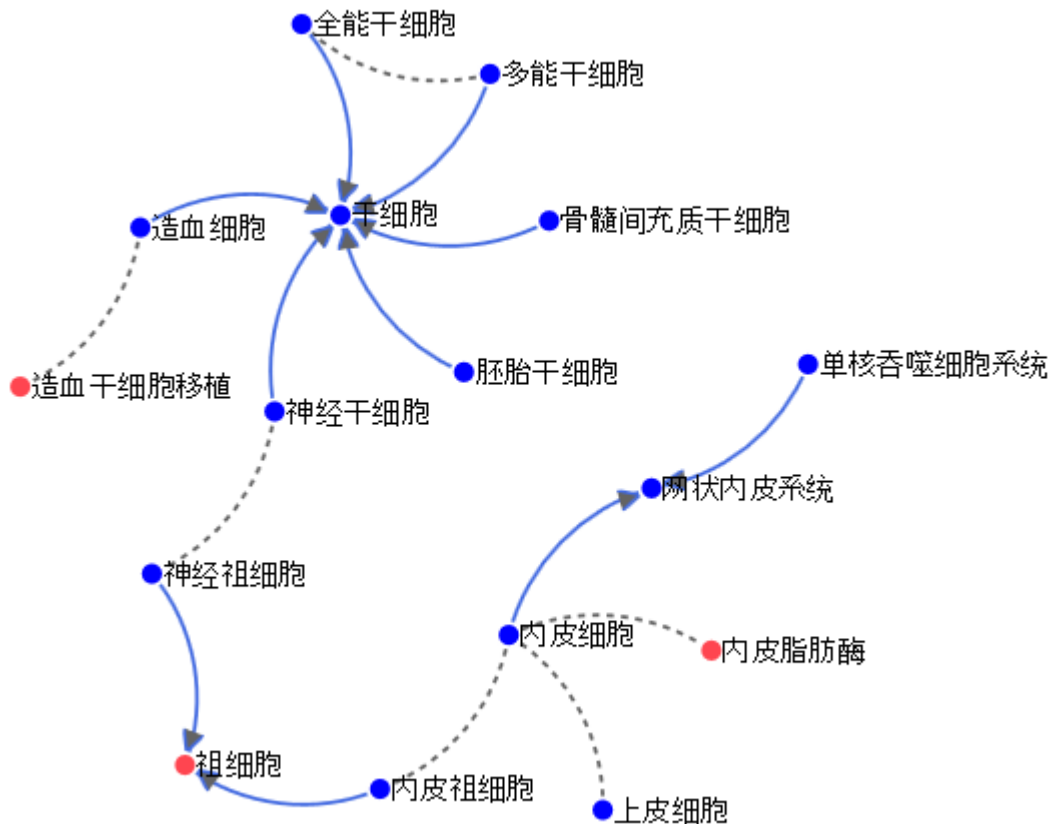
課題: 専門用語の抽出が不十分

Concept Map Searchの語のネットワーク表示

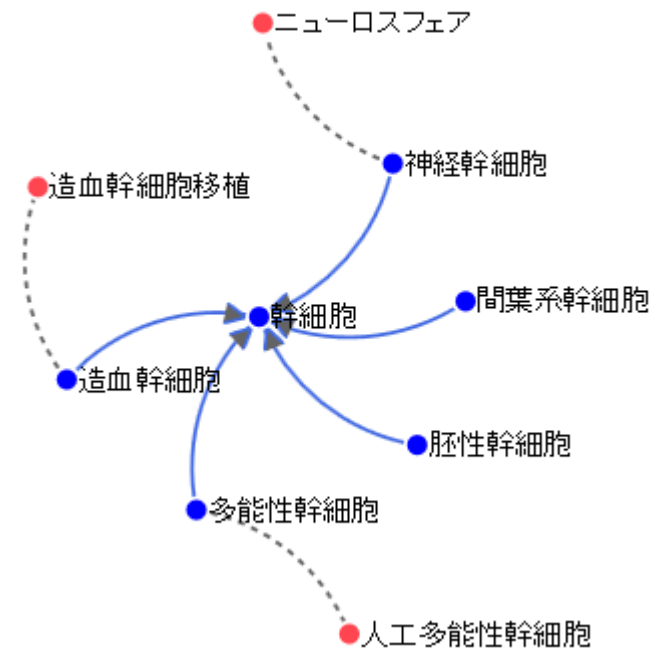
WIPO Pearl¹⁰のConcept Map Search

<http://www.wipo.int/wipopearl/search/home.html>

胚胎干细胞 (pēitāigànxìbāo) 中国語



ES細胞 (ESさいぼう) 日本語



- キーワードから公報へリンクされている
- インタラクティブにリンク操作できる(適合性に有利)
- 収録語数・収録公報数が少ない

「iPS」周辺の中国語分詞結果のネットワーク

媒介中心性(Betweenness Centrality)でカラー(サイズ)マッピング

Control Panel

Network Style Select

Network	Nodes	Edges
Sheet1		
Sheet1	895(0)	1998(1)

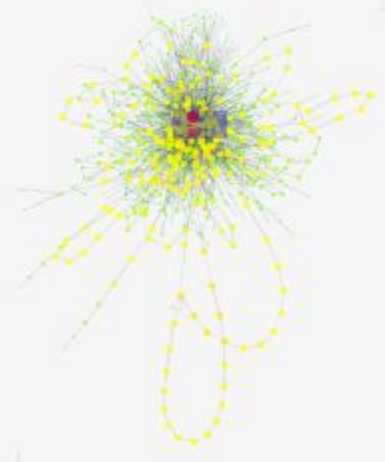
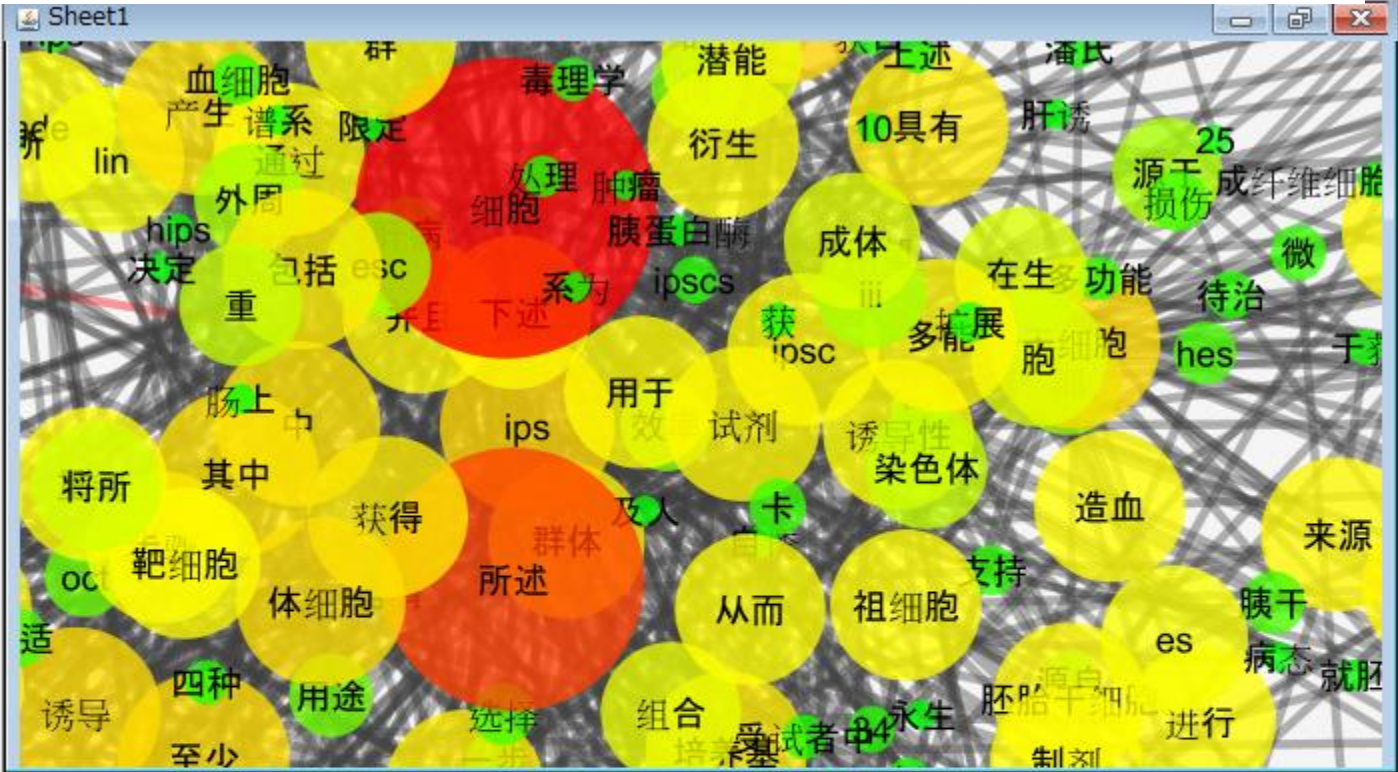


Table Panel

shared ...	name	△ BetweennessCentrality	Average...	Clusteri...	Closen...	EdgeC...	Indegre
细胞	细胞	0.33300983	3.21121252	0.01073537	0.31140885	194	8
所述	所述	0.25042431	3.38200782	0.02019044	0.29568234	114	5
方法	方法	0.06638817	4.06910039	0.02255892	0.24575457	56	4
包含	包含	0.06197898	3.78226858	0.04761905	0.26439159	37	1
一种	一种	0.06158151	4.04302477	0.01748252	0.24733957	67	3
ips	ips	0.05750401	4.10169492	0.03764706	0.24380165	53	4

Node Table Edge Table Network Table

非類似度(距離)による公報のネットワーク

Threshold: 0.45以下を表示

類似度計算: Cosine類似度

Cytoscape3.2.0

Session: D:\Data\Doc\情報\アジア特許情報\INFOPRO2014\IPS2\Cytoscape\CN286閾値0.45-n0.cys

File Edit View Select Layout Apps Tools Help

Control Panel

Network Style Select

Network	Nodes	Edges
Sheet1	225(1)	3971(4)

Networkデータ(一部抜粋)

Doc1	Doc2	非類似度(距離)
CN101356270	CN101617043	0.340
CN101356270	CN101855350	0.379
CN101356270	CN102822345	0.351
CN101356270	CN103562376	0.446

Sheet1

注目特許

CNIPRの相似性検索

上から3,4位

Table Panel

属性データ

shared na...	name	CN-TI	NPA	IC
CN101356270	CN101356...	核重新編因子	KYOTO NATIONAL ...	A61K-035/12 A61L-0

Node Table Edge Table Network Table

まとめ

中国語KWを用いた特許情報解析に関して、下記の一連の流れを検討した。

- ①キーワードのネットワーク分析による重要語(特徴語)抽出
 - ②キーワードを用いて計算した公報の非類似率によるネットワーク分析
 - ③適合率を重視した特許調査への応用
- ・重要なKWを選択することで中国特許調査の調査精度、特に適合率の向上に有用である。
 - ・ターゲット公報を解析した発明の特徴を現す重要KW抽出が適合率向上のポイントである。
 - ・重要KW抽出にネットワーク分析による媒介中心性が有用である。
 - ・公報の非類似度(距離)を使用してネットワーク表示が可能である。
 - ・注目公報の近くにある公報から確認することで適合性重視の調査ができる。

今後の応用

- ・中/日機械翻訳の翻訳品質のテキストマイニングを利用した評価
- ・中/日機械翻訳の日本語での検索の再現率、適合率を上げる方法
- ・テキストマイニングを利用した動向調査

謝辞

最後に、本報告は2014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す。研究会のメンバーの皆様には様々な協力をしていただきました。特に昨年度の研究会メンバーである山村健一氏には貴重なアドバイスをいただきました。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。

参考文献

- [1] 安藤 俊幸ら. “中国語キーワードによる中国特許情報解析” 第10回情報プロフェッショナルシンポジウム
- [2] IKAnalyzerNet.
<http://www.piaoyi.org/c-sharp/IKAnalyzerNet.html> accessed 2014.10.25
- [3] ICTCLAS.<http://ictclas.nlpir.org/> accessed 2014.10.25
- [4] saezuri lite.
<http://www.vector.co.jp/soft/winnt/prog/se495669.html> accessed 2014.10.25
- [5] Cytoscape <http://www.cytoscape.org/> accessed 2014.10.25
- [6] 知的財産情報検索委員会第2小委員会. “中国特許調査に関する研究”. 知財管理 Vol. 63 No.12, (2013),1943-1957
- [7] RMeCab
<http://rmecab.jp/wiki/index.php?RMeCab> accessed 2014.10.25
- [8] 末吉正成ら. “テキストマイニングを行う”. Rではじめるビジネス統計分析. 翔泳社, 2014, p. 304-334.
- [9] 特許版事典検索システムCyclone
<http://cyclone.cl.cs.titech.ac.jp/> accessed 2014.10.25
- [10] WIPO Pearl
<http://www.wipo.int/wipopearl/search/home.html> accessed 2014.10.25
- [11] 語のネットワーク分析<http://www1.doshisha.ac.jp/~mjn/R/61/61.html> accessed 2014.10.25
- [12] Foster Provostら. “テキスト表現とテキストマイニング”. 戦略的データサイエンス入門. オライリー・ジャパン, 2014, p. 275-306.
- [13] 京都大学iPS細胞研究所. iPS細胞技術に関する基本特許1件が新たに日本で成立
<http://www.cira.kyoto-u.ac.jp/j/pressrelease/news/131220-203845.html> accessed 2014.11.25

文書間相互類似度計算(自作VB.Netプログラム)

VB 2008



推奨

- ①非類似度(距離)マトリックス計算(2次元)
 - ②類似度ソート用(1次元) ③統計出力
- 特徴

- ・全文書の特徴語、重要度をメモリ上に保持
- ・文書間の共通語の抽出にハッシュを使用
- ・正規表現によるノイズ除去機能

文書間相互類似度の組み合わせ数

$$\frac{n \times (n-1)}{2} \quad \text{1000件の文書場合} \quad \frac{1000 \times 999}{2} = 499500$$

推奨 →

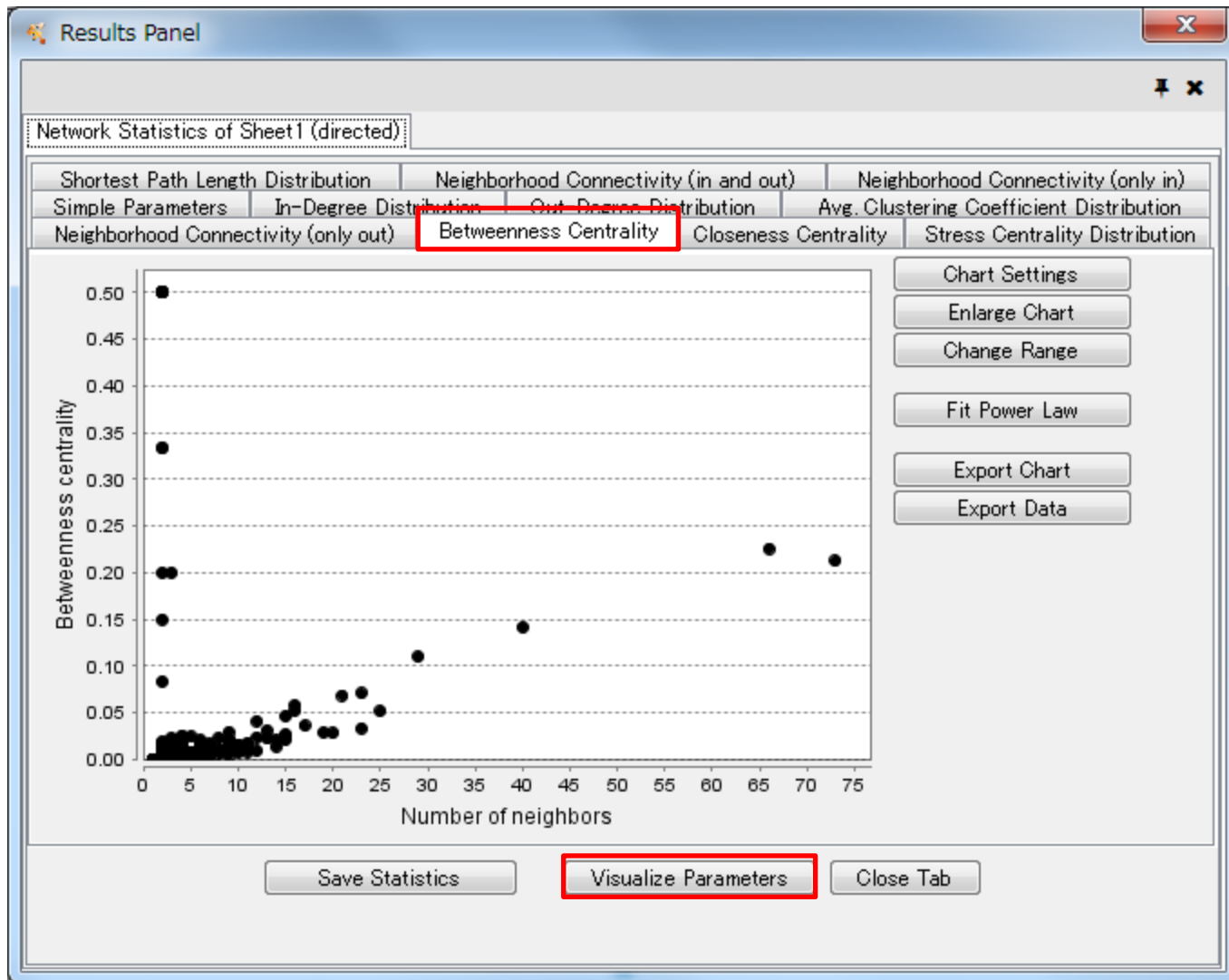
重み付け手法と類似度計算方法

	重み付け	類似度計算方法
1	2値	余弦(Cosine)係数
2	2値	ダイス(Dice)係数
3	2値	ジャカルル(Jaccard)係数
4	2値	重複(Overlap)係数
5	2値	単純一致c/a
6	2値	単純一致c/b
7	重み	余弦(Cosine)係数
8	重み	ダイス(Dice)係数
9	重み	ジャカルル(Jaccard)係数
10	重み	単純重み付き

解析対象

1. 一种体细胞的核重新编程因子，该因子含有下述3种基因：Oct家族基因、K1f家族基因和Myc家族基因的各基因的产物。
2. 权利要求1中记载的因子，其含有下述3种基因：Oct3/4、K1f4和c-Myc的各基因的产物。
3. 权利要求1或2中记载的因子，其还含有下述基因：Sox家族基因的基因产物。
4. 权利要求3中记载的因子，其含有Sox2的基因产物。
5. 权利要求1至4中任意一项中记载的因子，其含有细胞因子，该细胞因子和Myc家族基因的基因产物共同存在、或者代替Myc家族基因的基因产物。6. 权利要求5中记载的因子，所述细胞因子为bFGF和/或SCF。
7. 权利要求1至6任意一项中记载的因子，其还含有下述的基因：TERT基因的基因产物。
8. 权利要求1至7任意一项中记载的因子，其还含有选自于由下述的基因构成的组中的1种以上的基因的基因产物：SV40大T抗原、HPV16 E6、HPV16 E7和Bmil。
9. 权利要求1至8任意一项中记载的因子，其还含有选自于由下述的基因构成的组中的1种以上的基因的基因产物：Fbx15、Nanog、ERas、ECAT15-2、Tc11和 β -连环素。
10. 权利要求1至9任意一项中记载的因子，其还含有选自于由下述的基因构成的组中的1种以上的基因的基因产物：ECAT1、Esg1、Dnmt3L、ECAT8、Gdf3、Sox15、ECAT15-1、Fth117、Sa114、Rex1、UTF1、Ste11a、Stat3、和Grb2。
11. 通过体细胞的核重新编程制备诱导式多能性干细胞的方法，其包括使权利要求1至10任意一项中记载的核重新编程因子与该体细胞接触的步骤。
12. 权利要求11中记载的方法，所述体细胞为人细胞。13. 由权利要求11或12中记载的方法获得的诱导式多能性干细胞。
14. 通过诱导权利要求13中记载的诱导式多能性干细胞分化而得到的体细胞。
15. 一种改善细胞的分化能力和/或增殖能力的方法，该方法包括使权利要求1至10任意一项中记载的核重新编程因子与细胞接触的步骤。
16. 权利要求15中记载的方法，所述细胞为人细胞。
17. 由权利要求15或16中记载的方法获得的细胞。
18. 通过诱导权利要求17中记载的细胞分化而得到的体细胞。

ネットワーク分析方法(媒介中心性Betweenness Centrality)



ネットワーク分析結果の可視化設定

(媒介中心性Betweenness Centrality)

