

# 日本語処理技術の研究開発に おける特許情報の可能性

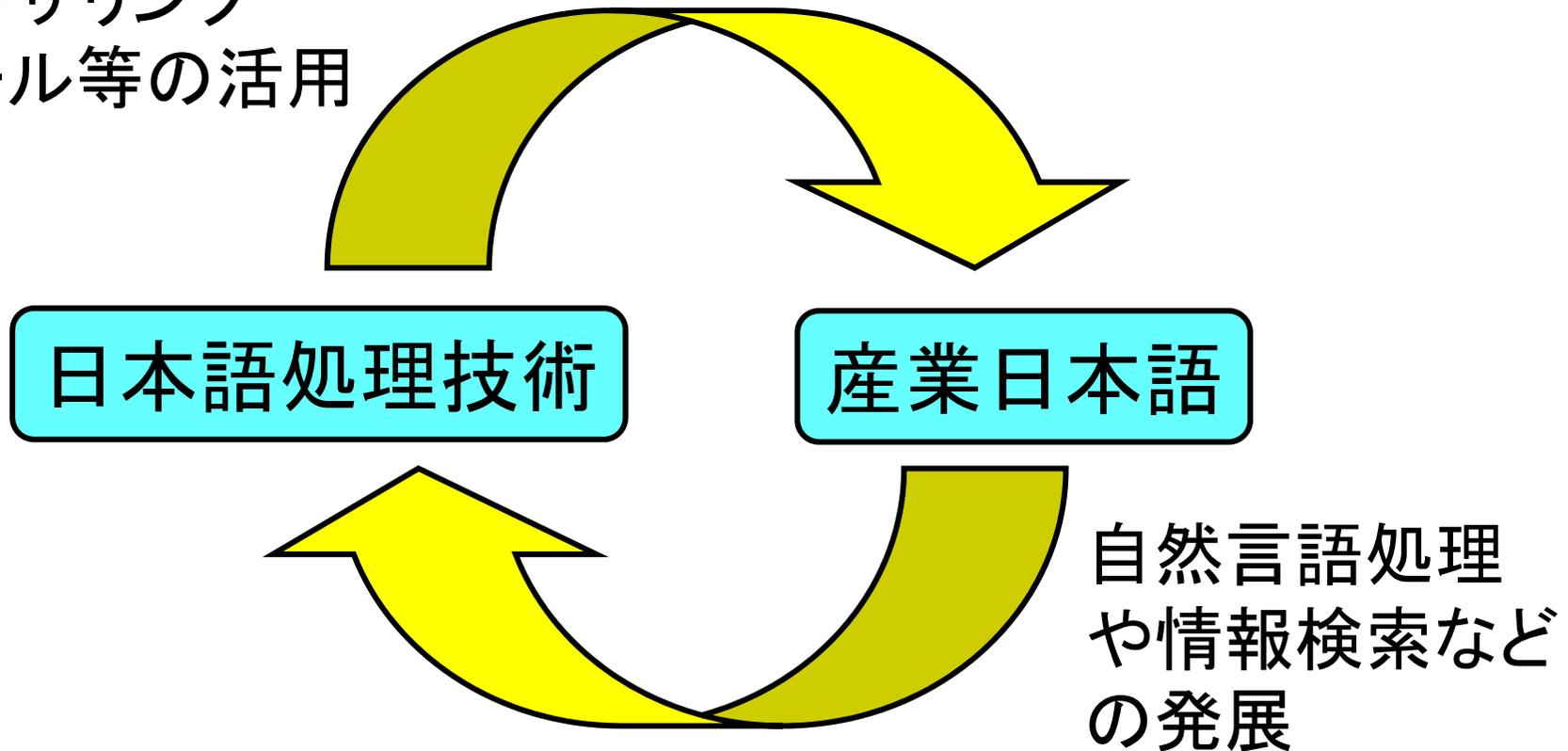
藤井 敦

東京工業大学

2010.02.24

# 本講演の背景：建設的な循環

オーサリング  
ツール等の活用



産業用文書としての**特許情報**に焦点を当てて、  
関連する日本語処理技術について紹介する

# 特許情報と学術研究の関係

- 情報検索
  - 要求に適う文書の検索
- 自然言語処理
  - 機械翻訳, 用語抽出
- 人工知能
  - 知識発見, マイニング
- データ工学
  - 構造化データの管理等

## チャレンジ

- 技術動向を知る
- 権利をつぶす
- 文書の構造を使う
- 玄人志向 vs 素人志向

- 特許情報は研究素材のデータとして有望
- 特許情報特有のチャレンジがある

# 特許・情報フェア & コンファレンス

## 2009年11月

- 情報検索
  - 自然文検索, 類似文書検索, 関連語抽出は当然
- 分析・可視化
  - 2次元パテントマップ
  - 特許どうしの引用関係
  - 請求項の従属関係
- 明細書作成・校正支援
  - 定型表現や用語説明等の再利用
  - 請求項における従属関係の検査
- 機械翻訳
- 文書管理, 知識共有, セキュリティ等

自然言語処理や情報検索の  
先端技術が応用されている

# 私と特許情報の関わり

- 多言語横断特許検索サービス(1998~2003)
  - 日本特許情報機構, (株)パトリスの受託研究

- NTCIRプロジェクト(2001~)
  - 特許情報処理用テストコレクションの構築
- NEDOプロジェクト(2005~2008)
  - 特許情報を用いた辞典検索システムの構築

本日の内容



- 情報検索や自然言語処理を評価するためのベンチマーク(テストコレクション)を構築する
  - 参加チームの協調と競争を通して構築する
- 1年半を周期として, 現在8回目(NTCIR-8)
  - 成果報告会:2010年6月15-18日@NII
- 様々なタスクがある
  - 言語横断検索, 要約, 質問応答, 意見分析など
  - 3回目(NTCIR-3)から特許検索を開始

# 特許関連タスクの歴史

- NTCIR-3 既存の検索問題を特許で実践
  - NTCIR-4 現場の特許調査に挑戦
  - NTCIR-5 大規模に展開
  - NTCIR-6 検索と分類を総括
  - NTCIR-7 新領域を開拓
  - NTCIR-8 さらに展開
- } 翻訳・マイニング

検索

分析

分類

翻訳

マイニング

## 最新のデータセット

- 日本公開特許公報15年分
- 米国特許公報15年分

研究目的には無償配布

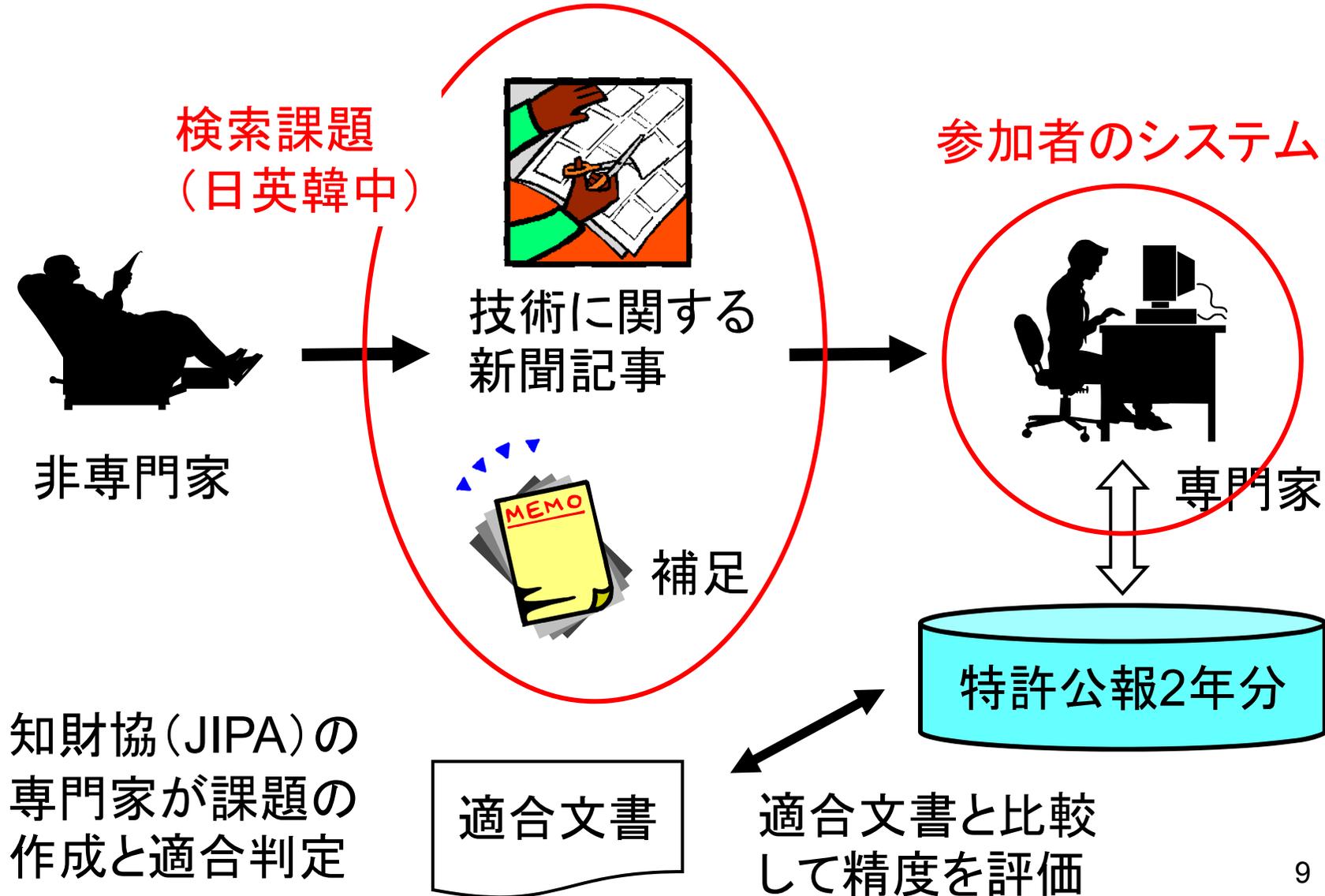
# NTCIR-3 (2001~2002年)

既存の検索問題を  
特許で実践

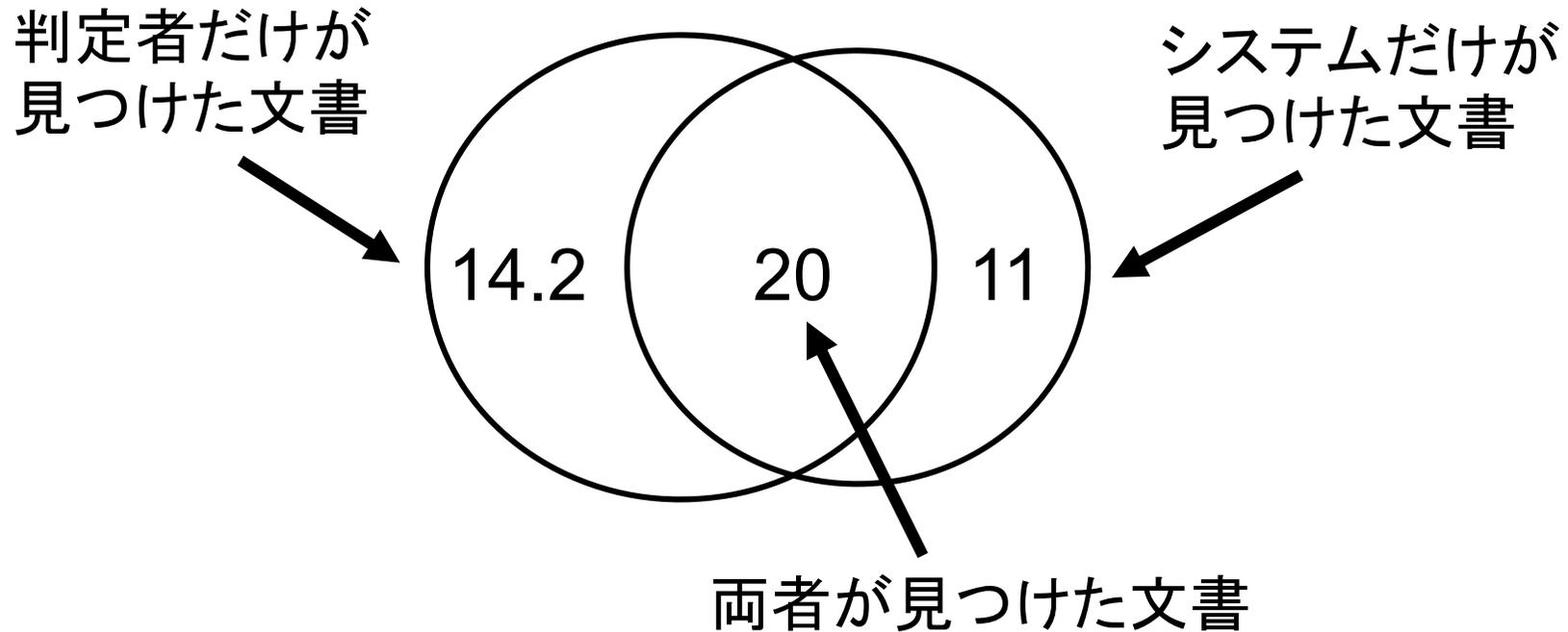
- 「技術動向調査」を模倣
  - 新聞記事を用いて特許を検索
  - 課題31件, 公開公報2年分(1998~1999年)
  - 知財協(JIPA)の委員による適合判定
    - JIPA委員による事前調査
    - プーリング

参加システムが検索した文書(上位)の  
集合だけを対象に適合判定を行う手法

# 技術動向調査タスクの概要



# 適合文書の内訳：31課題の平均



- 各システム(合計36)から上位100文書ずつ集めると, JIPAの判定者が見つけた正解の約80%を網羅できた

# NTCIR-4（2003～2004年）

## 現場の特許調査に挑戦

- 「無効資料調査」を模倣
  - 既存の請求項に類似した特許を検索
  - 課題101件，公開公報5年分（1993～1997年）
    - 34課題：JIPAの委員による適合判定
    - 67課題：特許庁の引例だけが適合文書（プーリングなし）
- 「特許マップ」の自動生成
  - 技術テーマに関する2次元マトリクスを生成
  - NTCIR-3の6課題を対象
  - JIPAの委員による添削（定量的な評価なし）

# 無効資料調査タスク

- 目的

- 請求項で主張されている権利を無効化する
  - 特許庁審査官が行う実体調査
  - 他者の権利を無効化するための社内調査
- 請求項を入力とした類似文書検索

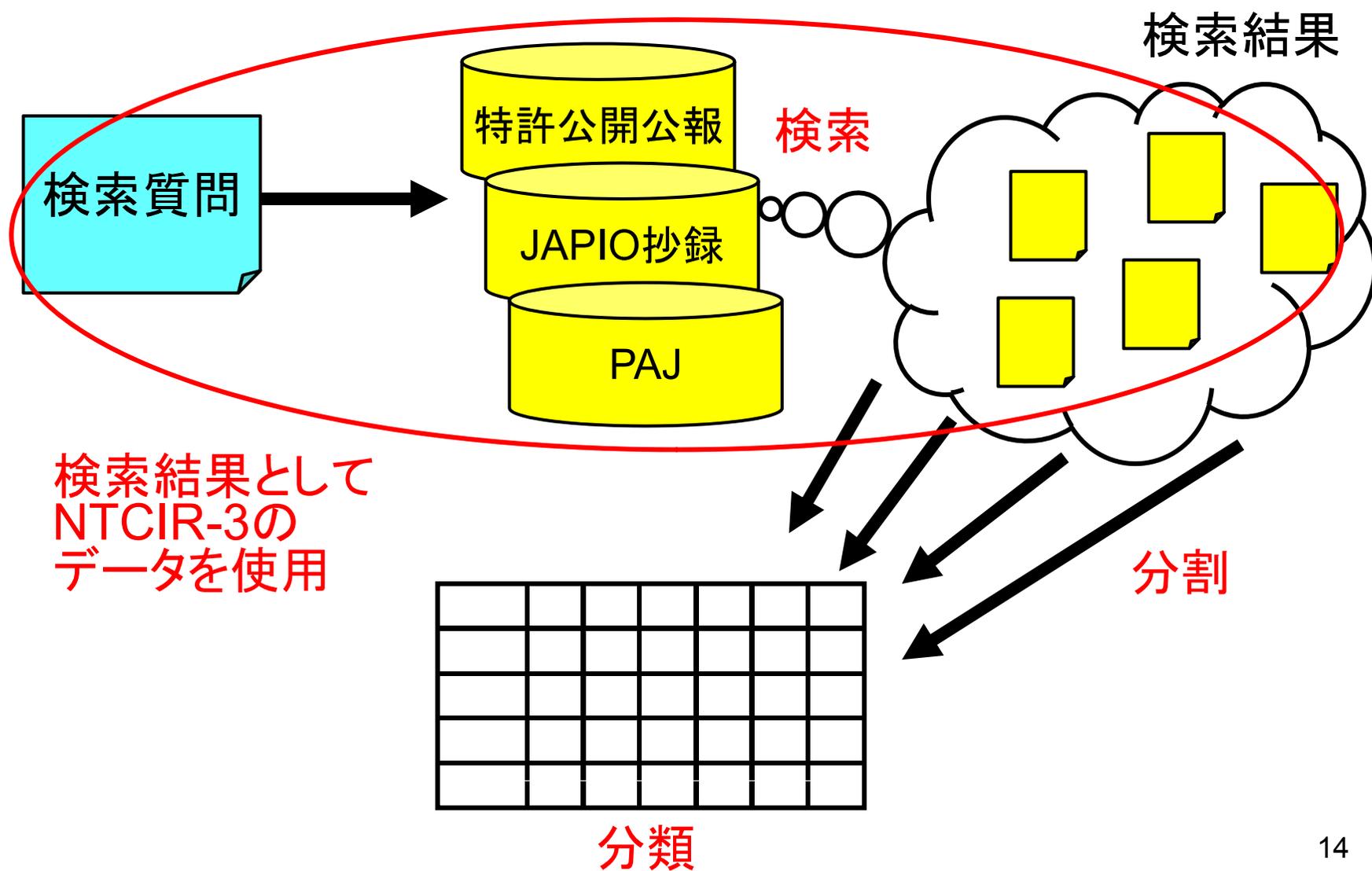
- 検索対象

- 公開特許公報
- 1993～1997年, 約170万文書, 40GB

# 特許マップ生成タスク

- 特許マップ
  - マクロ解析：全体の出願傾向を調査
  - ミクロ解析：テーマや分野を限定した調査
- NTCIRで目指した特許マップ
  - 分野や製品をテーマとして関連技術の分布を可視化する(ミクロ解析)
- 市販の特許マップ作成支援ソフト
  - 当たり前前の技術になってきた

# 特許マップ生成の概要



# 青色発光ダイオードの例（抜粋）

分類の軸は与えられている

課題

	結晶性	信頼性	長寿命	低抵抗	発光安定性	発光強度	発光効率
活性層構成			1998-145000 1998-233554				
基板構成	1998-256666				1998-022526		1998-107317
基板物性				1998-190146			
光吸収層構成			1998-022586				
中間層構成				1998-145006			

解決手段

分類項目と特許番号は参加チームが特定する

# NTCIR-5 (2004~2005年)

大規模に展開

- 文書検索タスク
  - 無効資料調査, 大量の課題
  - 課題1189, 公開公報10年分(1993~2002年)
  - 特許庁の引例だけが適合文書(プーリングなし)
- パッセージ検索タスク
  - 無効資料調査の根拠となる段落を検索
  - 課題356件(NTCIR-4の適合文書)
- 分類タスク
  - Fタームによる多観点の分類
  - 特許マップ生成の基礎技術

# NTCIR-7・8(2007～)

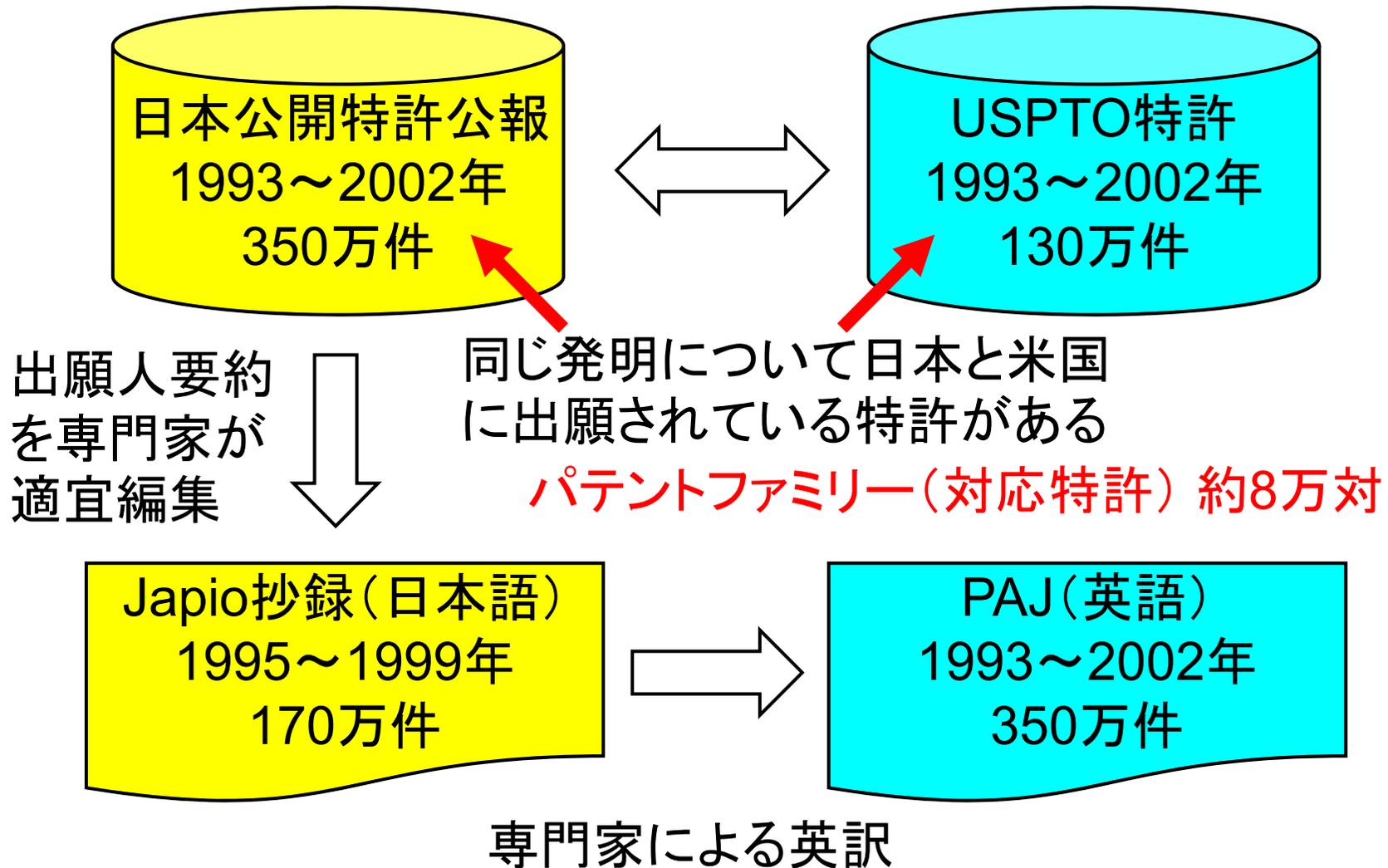
新領域を開拓

- 特許翻訳タスク
  - 機械翻訳
  - 言語横断検索
  - 翻訳の自動評価
- 特許マイニングタスク
  - 学術文献の分類
  - 技術動向マップの生成

# 特許翻訳タスク

- 動機
  - エンジンとデータが整備されつつある
- エンジン
  - 統計機械翻訳 (SMT) デコーダ
- データ
  - 日本公開公報とUSPTO特許 (1993～2002年) から抽出したパテントファミリー 8万対
  - 対訳文 200万対
    - 文翻訳の訓練データとテストデータとして使用可能

# NTCIRで配布している特許データ



※ NTCIR-8では1993～2007年(15年分)の特許公報を配布 19

# パテントファミリーの例

発明の名称: マイクロアクチュエータ



優先権主張番号でファミリーを特定することができる

Figure 1: Comparison of Japanese and US patent documents for a micro-actuator. The figure is divided into two main sections: Japan (left) and USA (right).

**Japan (Left):**

- (21) 出願番号: 特願平7-233220
- (22) 出願日: 平成7年(1995)8月24日
- (31) 優先権主張番号: 295, 127
- (32) 優先日: 1994年8月24日
- (33) 優先権主張国: 米国 (U.S.)
- (54) 【発明の名称】: マイクロアクチュエータ
- (57) 【要約】: 【課題】 所熱構造を備えるマイクロアクチュエータ... 【解決手段】 フローチャネルを介して運搬される流体... を制御する超小型バルブの形態をなすマイクロアクチュエータであり、サーマルアクチュエータによって選択的に駆動される熱膨張部材を有し、これが駆動されることによって熱エネルギーを生成する第1基板と、対向する第1、第2主要面を有する第2基板よりなる。第2基板が第1主要面で第1基板に取付けられる。第2の主要面は第2基板が支持体に取り付けられと絶縁セルを固定し、これによってマイクロアクチュエータの熱容量を減少させ、第1基板を支持体から熱断開する。

**USA (Right):**

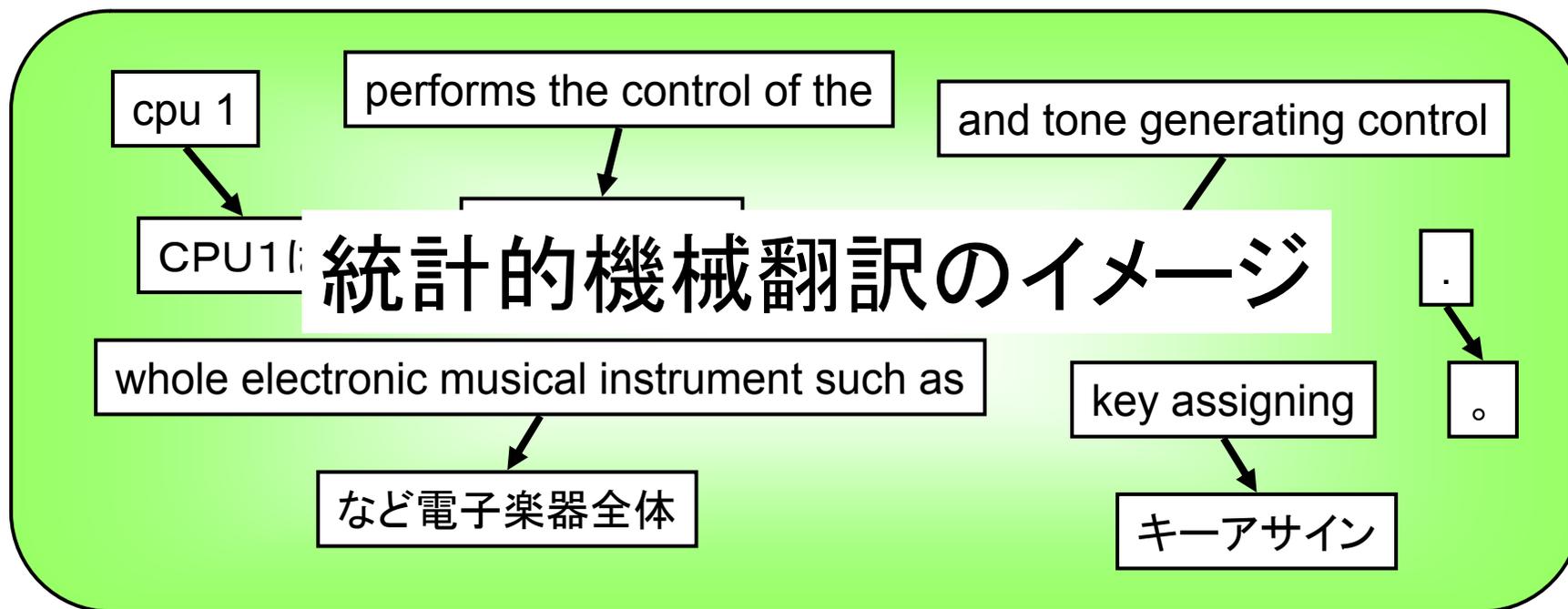
- (11) Patent Number: 5,509,279
- (45) Date of Patent: Jun. 25, 1996
- (54) THERMAL ISOLATION STRUCTURES FOR MICROACTUATORS
- (75) Inventors: Christopher C. Beatty, Landerberg, et al.; James W. Baker, Elkton, Md.
- (73) Assignee: Hewlett-Packard Company, Palo Alto, Calif.
- (21) Appl. No.: 295,127
- (22) Filed: Aug. 24, 1994
- (51) Int. Cl. F16K 31/02; F03G 7/06
- (52) U.S. Cl. 251/11; 251/129.01; 251/368; 60/528; 60/529
- (58) Field of Search: 251/11; 251/129.01; 251/368; 60/528; 60/529
- (56) References Cited: U.S. PATENT DOCUMENTS: 4,198,416 O'Connor, et al. 8/1991; 5,199,191 Beatty, et al.
- (57) ABSTRACT: A microactuator preferably in the form of a microminor valve for controlling the flow of a fluid carried by a flow channel includes a first substrate having a thermally-actuated member selectively operated by a thermal actuator such that the first substrate thereby develops thermal energy, and a second substrate having opposed first and second major surfaces. The second substrate is attached to the first substrate at the first major surface. The second major surface defines an isolation cell for enclosing a volume when the second substrate is attached to the support to thereby reduce the thermal mass of the microactuator and to thermally isolate the first substrate from the support.
- 18 Claims, 16 Drawing Sheets

Figure 2: Technical drawings of the micro-actuator. (a) is a cross-sectional view showing a substrate with a channel and a valve. (b) is a top-down view of the valve mechanism. (c) is a side view of the valve mechanism. (d) is a cross-sectional view of the valve mechanism in a different state. Labels include 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100.

日英の対応する項目から単語やフレーズの単位で統計的な翻訳のモデルを学習することができる

原文: cpu 1 performs the control of the whole electronic musical instrument such as key assigning and tone generating control .

- ① ↓
- 単語やフレーズの単位で英語から日本語に翻訳する
  - 英日の対応は大量のテキストデータから学習しておく



- ② ↓
- 日本語として自然な語順に並べ替える
  - 日本語の語順も大量のテキストデータから学習しておく

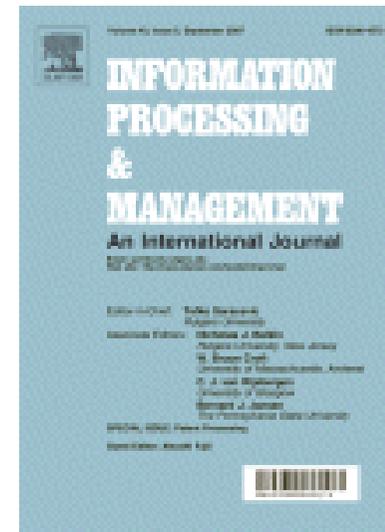
日本語訳: CPU1はキーアサイン、発音制御など電子楽器全体の制御を行う。 1

# 特許マイニングタスク

- 長期目標
  - 特許と論文を含む技術動向マップを自動生成
- 中期目標 (NTCIR-8)
  - 特許公報から要素技術や効果を抽出
- 短期目標 (NTCIR-7)
  - 論文をIPCに基づいて分類

# NTCIRに関連した活動

- 統計的機械翻訳の講習会
- ワークショップ
  - ACM SIGIR 2000 : Patent Retrieval
  - ACL 2003 : Patent Corpus Processing
- 論文誌特集号
  - Information Processing & Management
    - Patent Processing 特集号
    - Vol.43, No.5, 2007
    - 検索, 分類, マイニング等



# NEDOのプロジェクト(2005～2008)

- 特許情報を用語辞典として使う
- Cyclone検索エンジンの技術を応用
  - 抽出対象をWebページから公開特許公報へ
- 見出し語数：約190万語

<http://cyclone.cl.cs.titech.ac.jp/>  
で公開中

# なぜ特許情報から用語辞典を作るのか？

Webになく特許情報には存在する用語があるから

ジルコニウムジクロリド, 重合体成分, 感光性平版印刷版, 焼付定盤, 絶縁基体, 沃臭化銀乳剤, ハロゲン化銀乳剤, スルファモイル基, プラテンドラム, 塩基プレカーサー, エチレン性不飽和単量体, トラッキング誤差信号, 静電潜像保持体, スロットル弁開度, マゼンタカプラー

大抵はWebに掲載された特許が検索される

# 特許情報（公開公報）の構造

書誌事項

出願人要約

特許請求の範囲（請求項を記載）

請求の範囲を広げるために上位概念による抽象的な表現が使われる

発明の詳細な説明

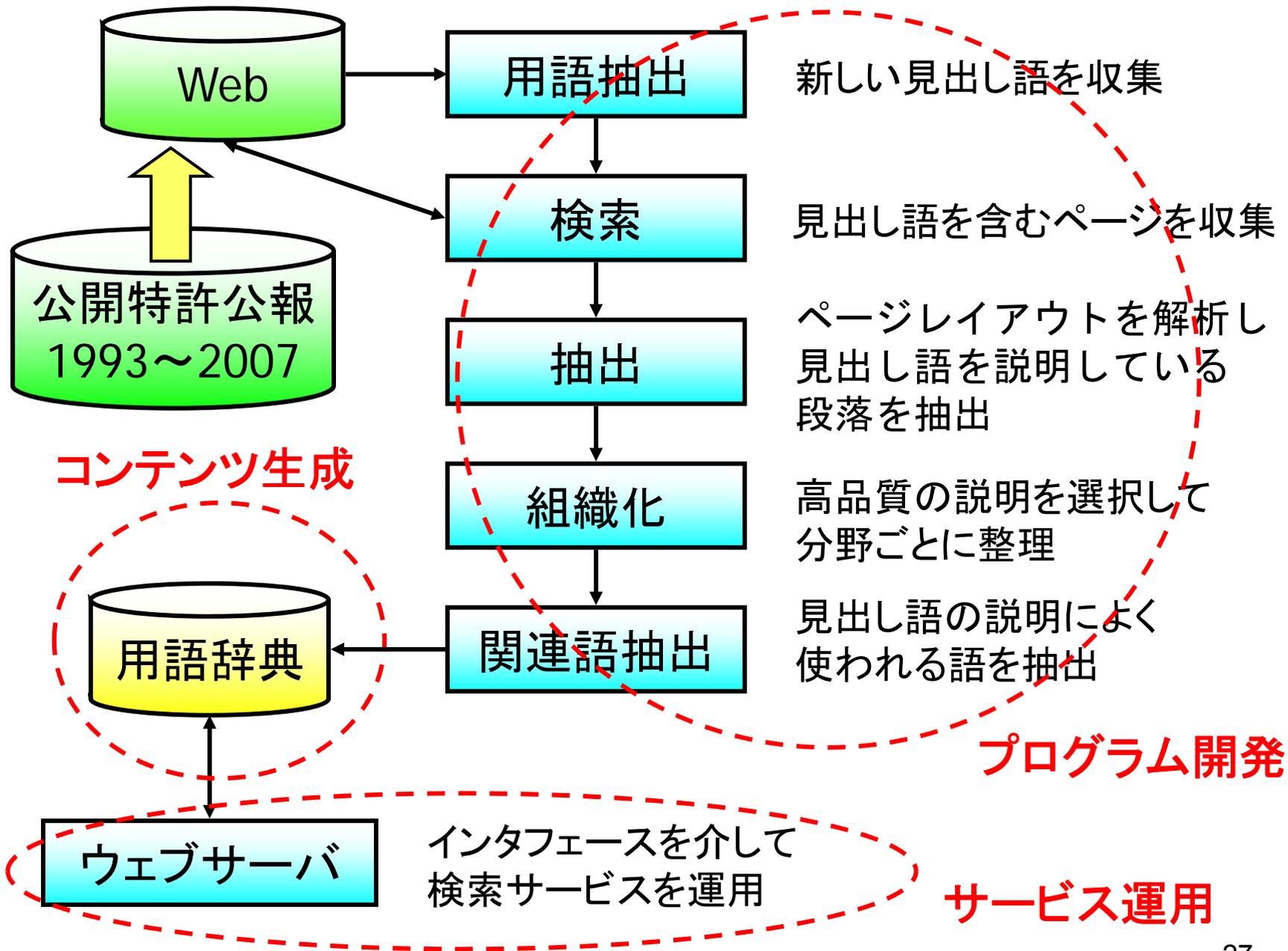
従来技術，課題，解決手段，実施例，発明の効果などを平易に記述する

図面の説明

日常言語  
と異なる

本研究の  
対象

ここから用語  
辞典を作る



# トラッキング誤差信号

検索語  検索 消去 関連語マップ 「トラッキング誤差信号」についてもっと調べたい

分野名  電気・電子  コンピュータ  機械  数学・物理  エネルギー  心理学  法律  プラント  スポーツ  衛生

関連語  **トラック**  **光ディスク**  反射光  光ビーム  光スポット  出力信号  対物レンズ  光ディスク装置  **トラッキング制御**  フォーカス誤差信号 >> つづ

複合語  前記トラッキング誤差信号  レンズ誤差信号生成手段  **トラッキング誤差信号TE**  トラッキング誤差信号検出手段  レンズ誤差信号検出装置  上記トラッキング誤差信号 >> つづ

「トラッキング誤差信号」の説明を検索しました  
分野名を選択すると、その分野の説明一覧が精

**関連語：光ディスク，対物レンズ，など**

**用語説明**

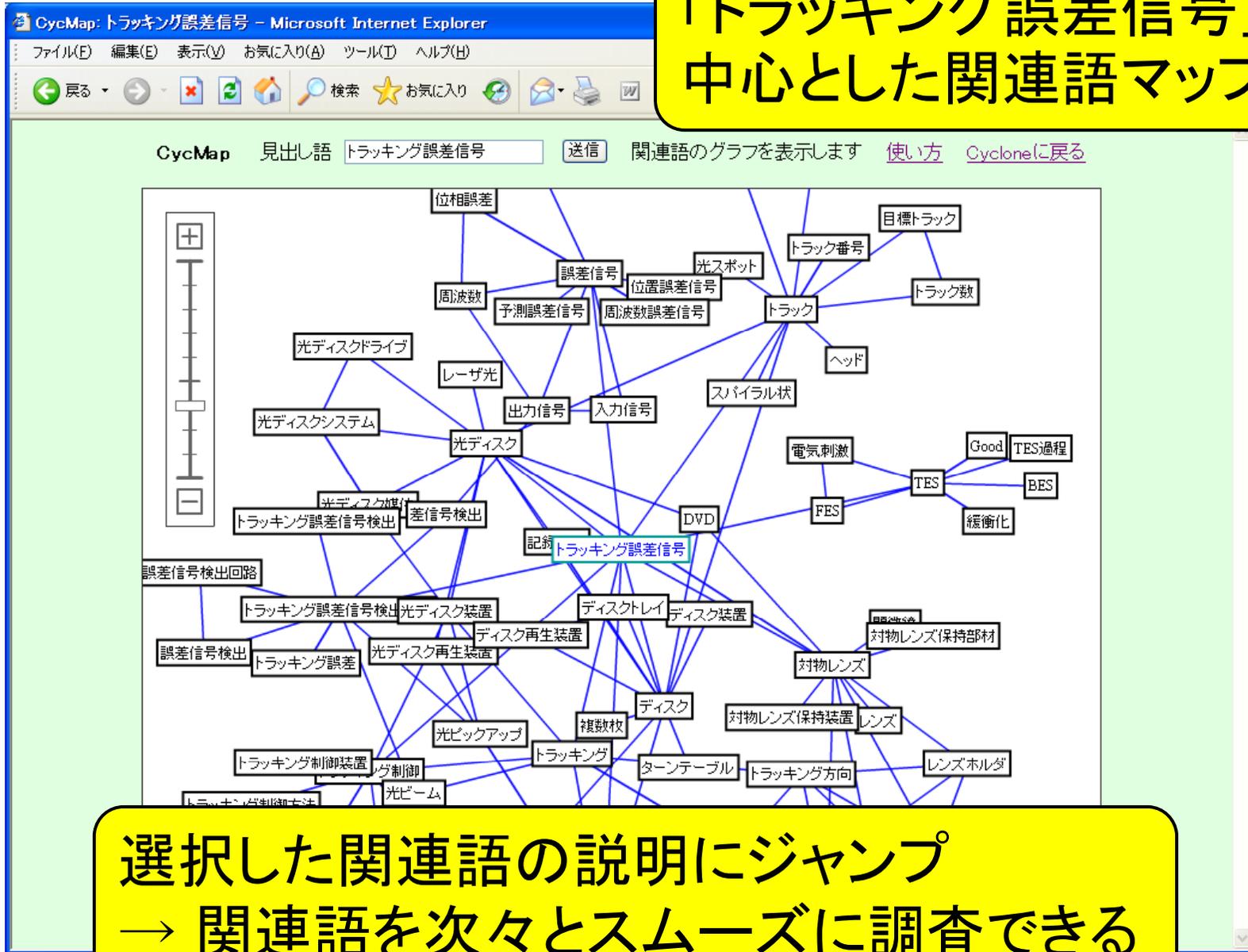
**光ディスク装置 [電気・電子]**

ブレーキ回路は、対物レンズがトラックを横断するとき、フォワード方向とリバース方向とでEFM信号と**トラッキング誤差信号**との位相関係が180度反転することを利用して**トラッキング誤差信号**の不要な部分をカットしてトラッキングアクチュエータにブレーキをかけるためのブレーキ信号を発生する。ここで、EFM信号とは、光ディスクにおけるピットの有無を示す信号であり、データの読み取り信号に用いられる。**トラッキング誤差信号**とは、対物レンズが光ディスクの半径方向に移動する場合、記録されたピットの中心で0となり、ピットの中心からずれるに従って、値が大きくなる信号である。

**ディスク再生用のトラッキングサーボ回路 [コンピュータ]**

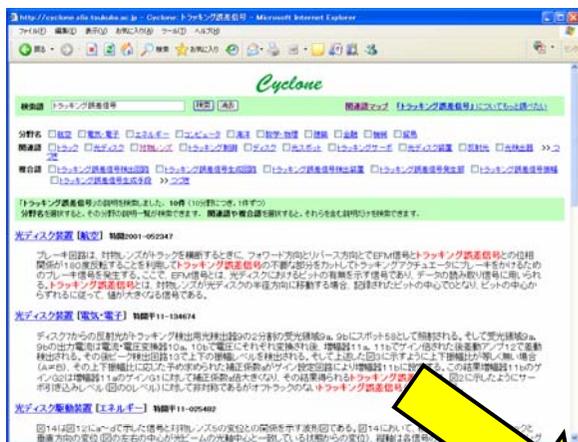
以下本発明の実施の形態について図面を参照して説明する。図1は本発明に係るディスク再生用のトラッキングサーボ回路の例を説明する図である。本図に示すトラッキングサーボ回路はCD又はMDに共通に使用され得るものであり、サイドスポット用の光ピックアップ検出器1と、光ピックアップ検出器の電流信号を電圧に変換し2つのサイドスポットの電圧を相互に減算して**トラッキング誤差信号**(TE)を形成する電流電圧変換及び減算回路2と、**トラッキング誤差信号**の位相補償を行う位相補償回路3と、トラッキングのループの開閉を行うスイッチ5と、**トラッキング誤差信号**を増幅する増幅器5と、レンズを有する光ピックアップ1のトラッキングを行うトラッキングアクチュエータ6と、位相補償回路3の出力である**トラッキング誤差信号**を入力してスイッチ5をオフからオンに制御し、且つ増幅器5にレンズをキックするレンズキック信号を出力してトラックジャンプを制御する制御部7とを具備する。

# 「トラッキング誤差信号」を中心とした関連語マップ



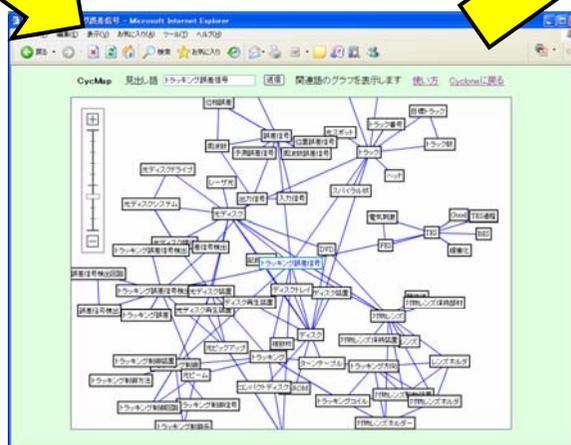
選択した関連語の説明にジャンプ  
→ 関連語を次々とスムーズに調査できる

# 関連語マップによる発想型検索



テキストによる説明  
(トラッキング誤差  
信号)

テキストによる説明  
(対物レンズ)



関連語マップ

# 辞典コンテンツを応用した特許検索

用語の意味を調べながら  
調査対象の分野に精通



関連語, 上位語, 下位語による発想支援

関連語による検索語の拡張

例) 有機EL → 発光層, ガラス基板

請求項中の抽象語を具体化

例) 移動体 → 電車, 飛行機

# 用語説明抽出の応用

- 明細書作成支援
  - 専門用語の定型的な説明文を再利用する
- 明細書の読解支援
  - 専門用語の説明文を除いて、発明の本質だけを読む

# Web版と特許版の比較：詳細度

- 特許版では計算式が出る
  - **オーバーラン**(空気の含有率)  
$$\{[(\text{一定容積のクリーム重量}) - (\text{一定容積のホイップ後のクリームの重量})] / \text{一定容積のホイップ後のクリームの重量}\} \times 100$$
- 特許版では調理時間や温度が具体的に出る
  - **半熟卵** → 98°Cで4分30秒
  - **温泉卵** → 68°Cで17分

# Web版と特許版の比較：意味

	Web版	特許版
コンプライアンス	法律遵守	変形のしやすさ
ドレッシング	料理	研磨
ブレイクスルー	前進	落ち込む現象
地下組織	グループ	根菜の部位

Webと特許の統合によって語義の網羅性が高まる

# まとめ

- 産業用文書である特許情報と日本語処理技術の関わりについて紹介した
- 自然言語処理と情報検索の融合を指向しながら研究を進めてきた
  - 特許検索タスクから特許翻訳タスクへ発展
  - 特許情報を用いた辞典的な検索の実現

# 謝辞：特許関連タスクの歴代オーガナイザ・ ご協力を頂いた皆様（敬称略）

- 神門 典子（NII）
- 岩山 真（日立製作所）
- 高野 明彦（NII）
- 難波 英嗣（広島市大）
- 内山 将夫（NICT）
- 山本 幹雄（筑波大）
- 宇津呂 武仁（筑波大）
- 橋本 泰一（東工大）
- 江原 暉将（山梨英和大）
- 越前谷 博（北海学園大）
- 下畑 さより（沖電気）
- 日本特許情報機構
- 特許庁
- 日本知的財産協会
- その他多くの皆様