

「企画」発想メモ 2009

平成21年9月17日(木)

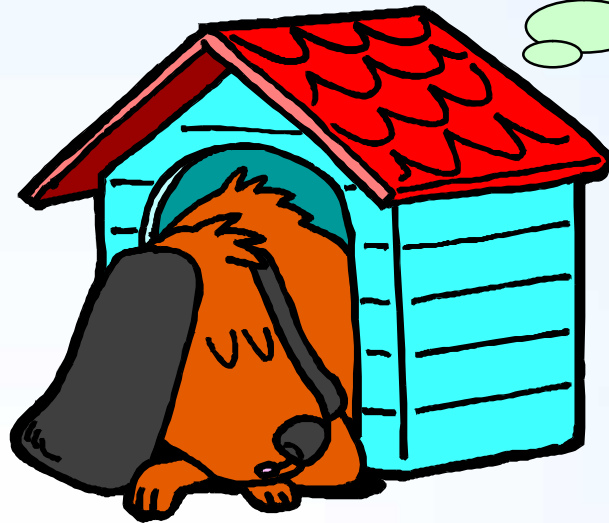
「図書系職員のための
アプリケーション開発講習会」

今期のまえだ企画

- 小川氏の「配架MAP」の社会科学研究所版を作る
 - 小川氏の成果を生かす
 - 講習会の職場還元
- 柏移管補助ツールのバッチ版を作る
 - 要望があったので、それに応える
- 「つっこみ用企画」(後述)をする
 - リスナーの受講生が**つっこむ**ための企画
 - ツールがいかに出来上がるか、半年かけてみてもらう
- 「東京大学OPAC Plus “言選Web”」で情報処理学会(来年3月)の発表を目指す
 - 自然言語処理・図書館資源の活用例
 - 情報基盤センターでの研究成果の還元

「配架MAP」社会科学研究所版

具体的にはこれから考えます



柏移管補助ツールの バッチ版を作る

具体的にはこれ
から考えます



つっこみメンバー用企画

- 今期の受講生には聴講生（つっこみ専門）が2名います。
- あまり退屈しなくて済むように、つっこみ用企画を立ち上げることにしました。

実もふたも、つっこみどころも ある話をめざして

- まえだがいままでに実例として作成したアプリケーション
 - 基本的に「実もふたもなし」
 - だいたい、30分から数時間の突貫開発で試作まで終わるから...
- ということで、**半年かけて、手の内をみせつつ**サンプルツールをつくります。

いつものお手軽開発の裏舞台

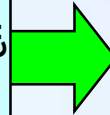
こんなツールを作れんかね



なんとかなるかも



この時点でシステムのパーツができていることも多いのです。



技術面でのネック(未経験技術)を調査

はやっ!



この企画いけそうな感じっす!



これでは、つっこみは難しい...

で、こうすることになりました

- ふだんどおりだと、企画の初期段階で
実も蓋もなく答えを出してしまいがち...
- このままだと、初心者向けの話にならないので、
 - 基礎技術から丁寧にサンプルプログラム
を作って説明します。
 - 今回ばかりはゴールできるか、自分でも
事前に調べないことにします。

つっこみ企画(案)

- ▶ WebCat, CiNii, OPACに関連語提示機能をつける
 - Yahoo! 関連語を使うつもり
 - <http://ferret-plus.com/ykwsearch>
- ▶ 次世代型の図書館システムかOPACを試す
 - Project Next-Lあたり
 - 次期次期の図書館情報システムのリプレイスのための勉強(気が早いかも?)

後で、聴講生と相談してみます。

「東京大学OPAC Plus “言選Web”」で 情報処理学会の発表(来年3月)を目指す

➤ 論文執筆の効能(と考えていること)

- システムの広報になる
- システムを全体的に見直せる
- システムのドキュメントにもなる
- 専門的な知識を詰め込める
- ドキュメント書きの訓練にもなる
- せっきくの情報基盤センター研究室との連携を生かせる

➤ 論文執筆や発表準備は勤務外として行ないます。

* https://mbc.dl.itc.u-tokyo.ac.jp/UT_OPAC_Plus_gensenweb からアクセス可

「東京大学OAPC Plus “言選Web”」リニューアルの序破急

- [序] 自然言語処理もできる図書館員をめざし勉強
- [破] 国会図書館件名標目と「言選Web」、統計的手法を使ったテキストマイニングを試す

研究者はだ
しのことに
チャレンジ

- 手法

- ◆ 非負値行列因子分解による特徴語抽出
- ◆ ニューラルネットワークを用いた件名標目形ナビゲーション
- ◆ R言語を用いた文書クラスタリング

- ネック

- ◆ 統計的な手法はシステム負荷が大きすぎ
- ◆ 結果がいいのか悪いのか、判断が難しい

- [急] 簡単なところで「東京大学OPAC Plus “言選Web”」と国会図書館件名標目を組み合わせることに

学術系の資源をフル活用

➤ 情報源

- 国内学術Webサイト(Yahoo! Web検索)
 - ◆ Web検索で使うキーワードを受け付けます。
- TSUBAKI
 - ◆ 「朝食を食べない子供の増加」といったキーワードを受け付けます。自然言語処理の研究者が開発。
- CiNII
 - ◆ 論文検索で使うキーワードを受け付けます

➤ 関連語提示

- 「情報源」+ 「言選Web」 → 独自理論
- 国会図書館件名標目
- 日本語WordNet

いまどきの技術を活用

- Web APIを活用
- Ajaxを活用
- オブジェクト指向を(多少は)考慮してプログラミング

国会図書館件名標目を扱う Perlモジュールも(ついでに)開発

➤ Perlモジュール”MARC::NDLSH”

- タブ区切り形式(最近は、XML形式もあり)で配布されている国会図書館件名標目をPerlから参照するためのモジュール(ライブラリ)
- 有向グラフとして、用語を扱えるインターフェイスとより簡便なインターフェイスの両方を用意

トップ画面

東京大学OPAC Plus "言選Web"

国内学術Webサイトから関連学術用語のリストを得た上で、
東京大学OPACもしくは雑誌記事索引(国会図書館 PORTA)にナビゲートする！

1. 調査したい日本語の専門用語（フレーズ）をいれて「実行」ボタンをクリック！
2. 国内学術Webサイトから調べた「関連語」とそれをキーワードにした[東京大学OPAC](#)・雑誌記事索引リンクを提示します。
- (1) 「国内学術Webサイト」以外に、「東京大学内Webサイト」や「限定しない」ことを選ぶこともできます
- (2) 情報源は[Yahoo! Web検索](#)以外に、[TSUBAKI](#)(情報爆発プロジェクト 検索エンジン基盤)、[CiNii](#)(NII論文情報ナビゲータ)を選ぶこともできます
4. 入力したフレーズとシソーラスの登録語が一致した場合、「国会図書館件名標目(NDLSH)と日本語WordNetを使用」にマッチした場合、その情報も提示します
5. 抽出したフレーズにシソーラスの登録語があった場合、「類似語」アイコン  も表示します。
(アイコンにマウスを乗せると、シソーラス

フレーズ入力

実行ボタンをクリック

クローン羊
(exam 羊, "集合知")

☐ 東京大学内のWebサイトから調査 ☒ 国内学術Webサイト(ac.jp)から調査 ☐ 限定せず
(Webサイトの限定は「情報源」がYahoo! (デフォルト) の場合のみ有効です)

●関連語探索の情報源を選択できます。

Yahoo! Web検索
Yahoo! Web検索
TSUBAKI (情報爆発プロジェクト 検索基盤)
CiNii (NII論文情報ナビゲータ)

情報源を選択

[修正記録](#)

可能性があります。

関連語提示画面(その1)

「クローン羊」の関連用語リストと 東京大学OPAC及び国会図書館PORTAへのリンク

[Windowを閉じる](#)

「OPAC」ボタンをクリックすることで、その用語についての東京大学内所蔵図書を、「雑誌記事」ボタンをクリックすることで雑誌記事索引（国会図書館PORTA）を調べることができます。

(国内学術サイトに限定して調査)



重要度順の関連用語リスト

シソーラスに登録されている語があった場合、 アイコンを表示。アイコンの上にカーソルを乗せるとシソーラスの類義語を（あれば）表示します。

-   [クローン羊](#)
-   [クローン](#)
-   [クローン人間](#)
-   [クローン牛](#)
-   [クローン技術](#)
-   [リー](#) 
-   [体細胞クローン羊](#)
-   [羊](#) 
-   [体細胞](#)

SIMアイコンに
マウスカーソルを
乗せると、類義語を表示

羊(家畜)/ 羅紗綿/ 綿羊/ 緬羊/ Sheep/ ヒツジ/ 家畜ヒツジ

関連語提示画面(その2)

「羊」の関連用語リストと 東京大学OPAC及び国会図書館PORTAへのリンク

[Windowを閉じる](#)

「OPAC」ボタンをクリックすることで、その用語についての東京大学内所蔵図書や、「雑誌記事」ボタンをクリックすることで雑誌記事索引（国会図書館PORTA）を調べることができます。

(国内学術サイトに限定して調査)

シソーラスでは「羊」について、以下のエントリーが見つかりました。

類義語: [羊\(家畜\)](#) / [羅紗綿](#) / [綿羊](#) / [領羊](#) / [Sheep](#) / [ヒツジ](#) / [家畜ヒツジ](#)
: [\[類義語に範囲を広げて関連語を再探索\]](#)

関連語: [羊毛](#)
: [\[関連語に範囲を広げて関連語を再探索\]](#)

上位語: [反芻目](#) / [牡羊](#) / [雄羊](#) / [ヒツジ](#)
: [\[上位語に範囲を広げて関連語を再探索\]](#)

下位語: [羊科](#)
: [\[下位語に範囲を広げて関連語を再探索\]](#)

入力したフレーズ
がシソーラスに登
録されていた場合
は、その「類義語」
等を提示

重要度順の関連用語リスト

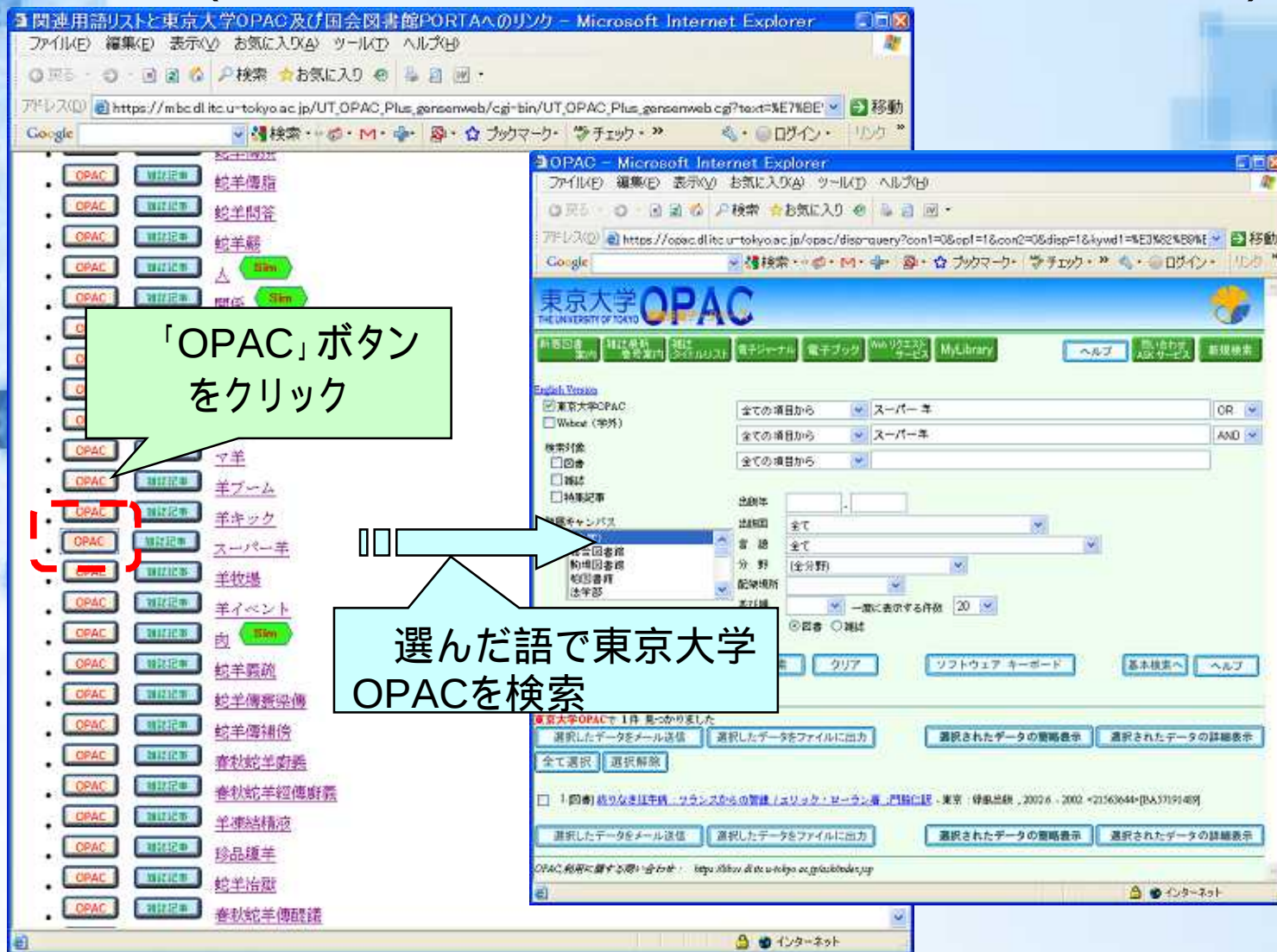
シソーラスに登録されている語があった場合

「下位語に範囲を広げて関連語を探索」など、
入力したフレーズからより範囲を広げた関連
語探索も可能

を（あれば）表示します。

[OPAC](#) [雑誌記事](#) [羊](#) [Shew](#)
[OPAC](#) [雑誌記事](#) [羊](#) [羊](#)

東京大学OPACへのリンク
(雑誌記事索引リンクもあり)



ゼロ件ヒット対策

➤ フレーズ検索と単語のAND検索の両方を行なう

● 東京大学OPACの場合

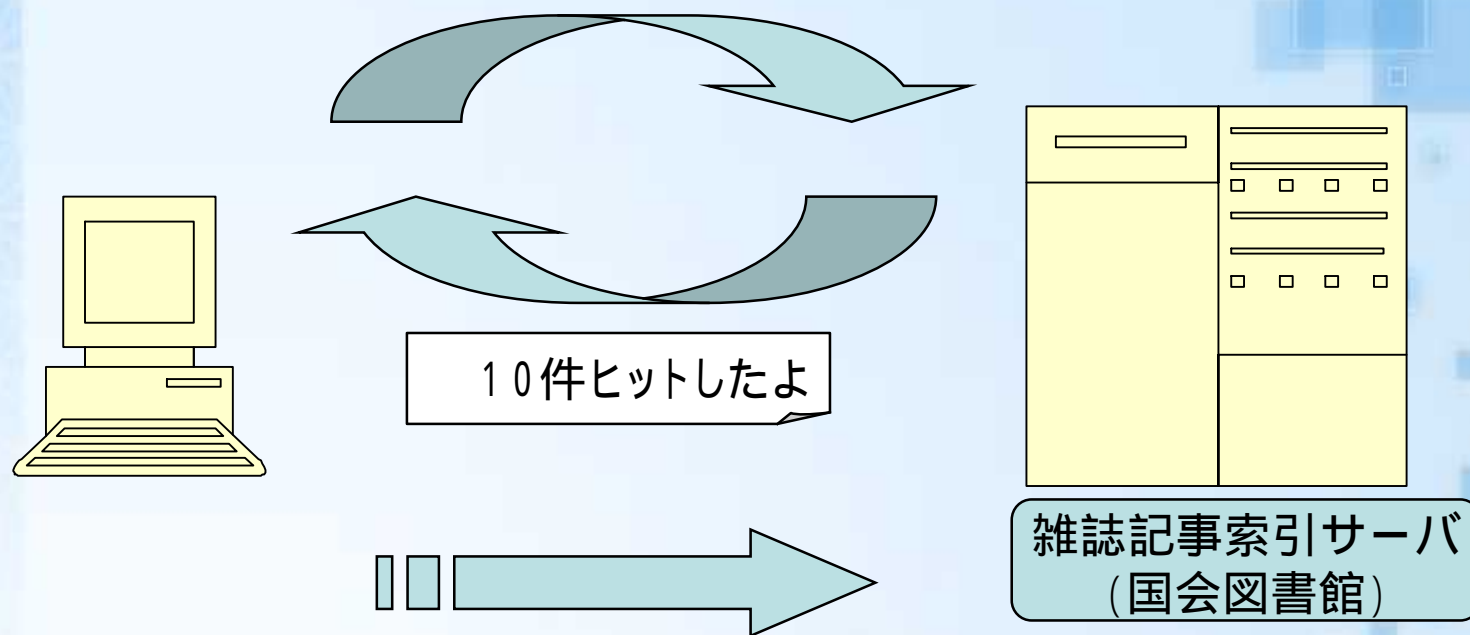
- ◆ フレーズ検索 OR 単語のAND検索
- ◆ 例 「東京特許許可局 OR (東京 AND 特許 AND 許可 AND 局)」

● 雑誌記事索引(国会図書館PORTA)の場合

- ◆ フレーズ検索でヒットするかチェック
- ◆ フレーズ検索で駄目な場合は、単語のAND検索

雑誌記事索引のゼロ件ヒット対策(1)

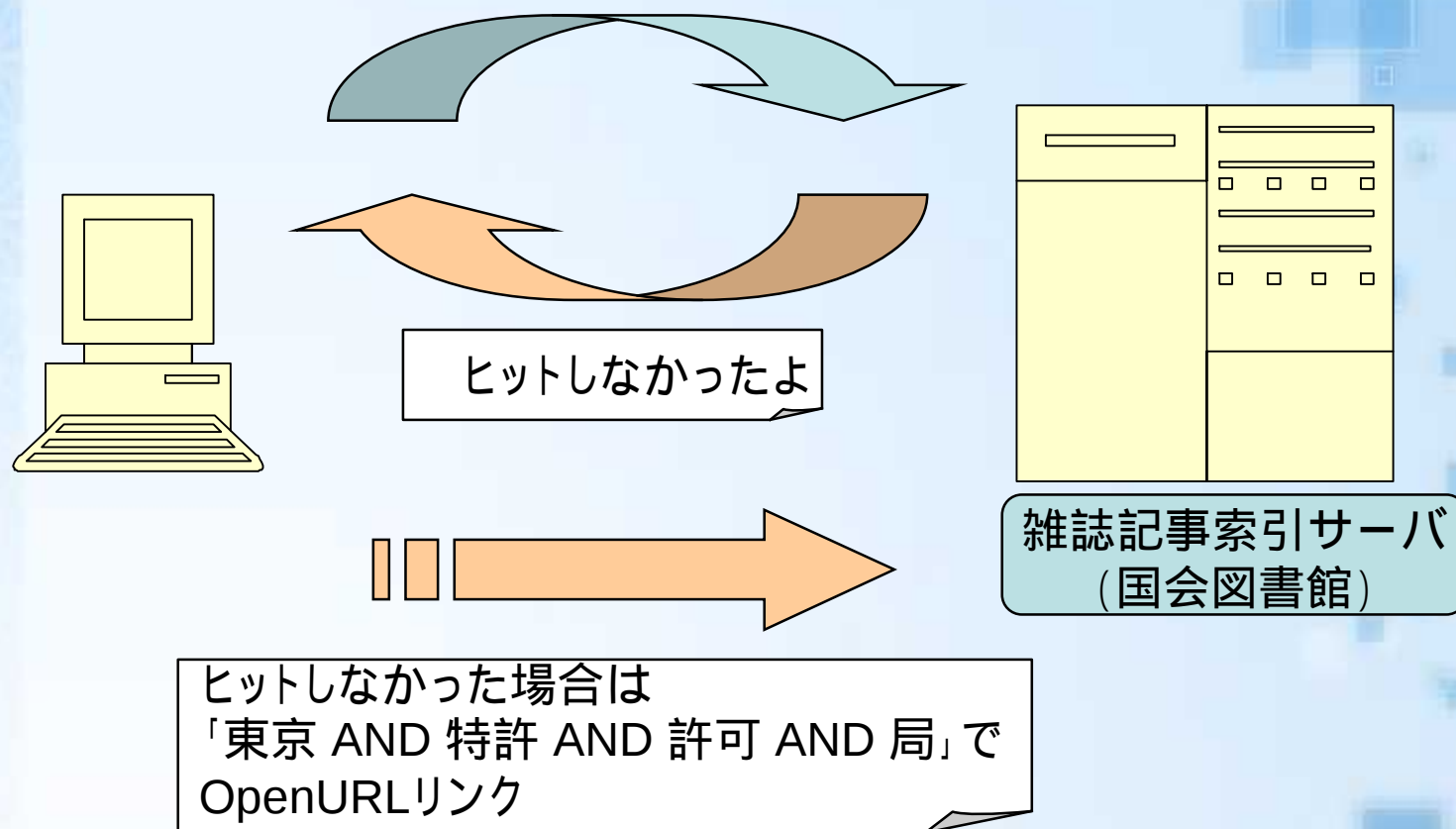
「東京特許許可局」でOpenSearch



ヒットしたことを確認して、
「東京特許許可局」でOpenURLリンク

雑誌記事索引のゼロ件ヒット対策(2)

「東京特許許可局」でOpenSearch



あとは、オフ会を

- 「いまどきの文献画像DB」のお話
- Perl&JavaScript講習
- 「OPACや電子ジャーナルをプログラムから使うには」(仮題)
- とか