

資源カード

「東京大学 OPAC」

東京大学 OPAC は、ブックマークレット、Firefox アドオン、CGI によるリンク、Web サービス化と組み合わせよう。

資源カード

「UT Article Search」(360 Search)

東京大学附属図書館の DB 横断検索。Web API を使うこともできる。

資源カード

「国会図書館 PORTA」

国会図書館の PORTA にある書誌や雑誌記事等を Web API を使って入手できる。

資源カード

「OCLC xISBN」

ISBN で検索すると、関連する資料（異版など）につけられた他の ISBN をタイトル情報込みで入手できる。

資源カード

「OCLC xISSN」

違版の ISSN の調査や、雑誌の変遷を調べることができる。

資源カード

「LC 件名標目の API」

Library of Congress Subject Headings API を使おう。

資源カード

「Google Book Search」

Book Viewability API を使って、Google Book Search が提供する「表紙画像」「書籍情報ページ」表紙のサムネイル画面」を自作アプリに組み込みできる

資源カード

「Amazon Web サービス」

Amazon 書籍の書誌情報、表紙画像、価格、カスタマーレビュー等を入手できる。

資源カード

「はてな Web サービス」

「はてな人力検索」の質問と回答を検索して取り出すことができます。

資源カード

「Web 検索」

Google や Yahoo!が Web 検索結果を XML で返すサービスを行なっています

資源カード

「地図情報」

Google や Yahoo!が地図情報を提供しています。

資源カード

「話題検索」

kizasi.jp の Web API を使うと、ブログで最近はやっている話題を検索できるみたい。

資源カード

「SNS 連携」

既存の SNS と連携させます。海外では FaceBook がよいみたいですが、日本ではどの SNS がよいのかな。

資源カード

「貸出記録」

個人情報がわからないように使うのは当然！
現状では使える環境にないので、とりあえず考えるだけ。

資源カード

「国会図書館件名標目表」

入手には国会図書館への申請が必要です。シソーラスとしても利用可能。

資源カード

「[国会図書館蔵書目録](#)」

理論カード

「協調フィルタリング」

アイテムに対するユーザの評価情報をもとに、アイテム同士の類似度や、ユーザ同士の類似度を導く。

理論カード

「クラスタリング」

データをグループにわけてしまう。ただし、どういう性質をもったグループにわけけるかはコントロールできない。

理論カード

「決定木」

与えられた条件から推論をおこなう。教師（学習）データをあたえて勉強させないと、動きません。

どういう筋道で推論をしたかがわかりやすく、中規模データでも結果がでるのがよいところ。

理論カード

「ニューラルネットワーク」

人間の神経をモデル化した手法。随時勉強をすることで、賢くなっていきます。

理論カード

「ベイズフィルタ」

電子メールのフィルタリングで大活躍。やはり随時勉強していくと賢くなります。でも、あまり複雑な条件設定ができないのが残念なところ。

理論カード

「SVM (Support Vector Machine)」

高性能な、分類器。ベクトル（配列）で表現できるデータであれば、ばっさばっさと2分します。事前の学習が必要。

理論カード

「遺伝的アルゴリズム」

最適なパラメータを求めるのに使用します。まさに自動思考錯誤マシン！ でも自作するなら模擬アニーリングのほうが楽かも。

理論カード

「K 近傍法」

条件が近いデータから推論を行ないます。たとえば、身長180cm なら、それに近い178cmの人と182cmの人の体重から、その体重を推定するとか？

理論カード

「非負値行列因子分解」

行列で表現できるデータから、特徴を取り出します。

理論カード

「N グラム」

文章を、N 文字（バイト）のパターンに着目し、解析する手法です。言語判定などでも利用。

理論カード

「固有表現抽出」

文章中から、人名、地名、等の名詞データを取りだします。

理論カード

「単語・文書行列」

文書とそれに含まれる単語頻度の表データです。文書同士の類似度や単語同士の類似度の計算の元になります。「連想検索」もこれが必要。

理論カード

「類似度計算」

とにかく、データをベクトル（配列）で表現しましょう。あとは、数学的な距離を使って類似度を計算できます。

理論カード

「 」

理論カード

「 」

理論カード

「 」

ツールカード

「GETA」

連想検索のほか、文書間のクラスタリングもできます。

ツールカード

「Algorithm::Cluster」

Perl のクラスタリングライブラリ。かなり使えます。

ツールカード

「C4.5」

決定木用のソフト。使い方はそう難しくないのですが、チャレンジの余地あります。

ツールカード

「TermExtract」

いわずとした、専門用語抽出 Perl モジュール。「言選 Web」のコアになります。

ツールカード

「Lingua::LanguageGuesser」

言語の判定（入力テキストがギリシャ語か、ドイツ語かフランス語か、など）を行なう Perl モジュールです。

ツールカード

「POLARIS」

入力テキストから、KeyGraph を簡単に生成できます。が、Windows のデスクトップアプリのため、他のプログラムとの連携は無理みたい。

ツールカード

「KH Coder」

有名なフリーのテキストマイニングソフト。Windows のデスクトップアプリです。

ツールカード

「CaboCha(南瓜)」

文章の係り受けを解析してくれます。固有表現の抽出機能つき。コマンドラインで動くので、組み込みも可能かも。

ツールカード

「vogoo」

これを使うと簡単に協調フィルタリングができてしまうみたい。PHP のライブラリ。

ツールカード

「Algorithm::NaiveBayes」

ベイズフィルタリングを行なう Perl モジュール。使ったことないので、使い勝手はなんとも。

ツールカード

「R 言語」

フリーの統計用言語。定番っぽいです。

ツールカード

「AI::FANN」

Perl 用のニューラルネットワークモジュール。みたかんじ、使いやすいそう

ツールカード

「[R のインストール](#)」

ツールカード

「[AI::FANN のインストール](#)」

ツールカード

「[R のインストール](#)」

ツールカード

「[AI::FANN のインストール](#)」