

株式  
会社

言語工学研究所

設立 1985

ホームページ <http://www.gengokk.co.jp>

E-メール [info@gengokk.co.jp](mailto:info@gengokk.co.jp)

# 事業内容

日本語の自然言語処理のシステムの開発  
OEM提供(API開示)

パッケージ販売

シソーラス 「デジタル類語辞典」

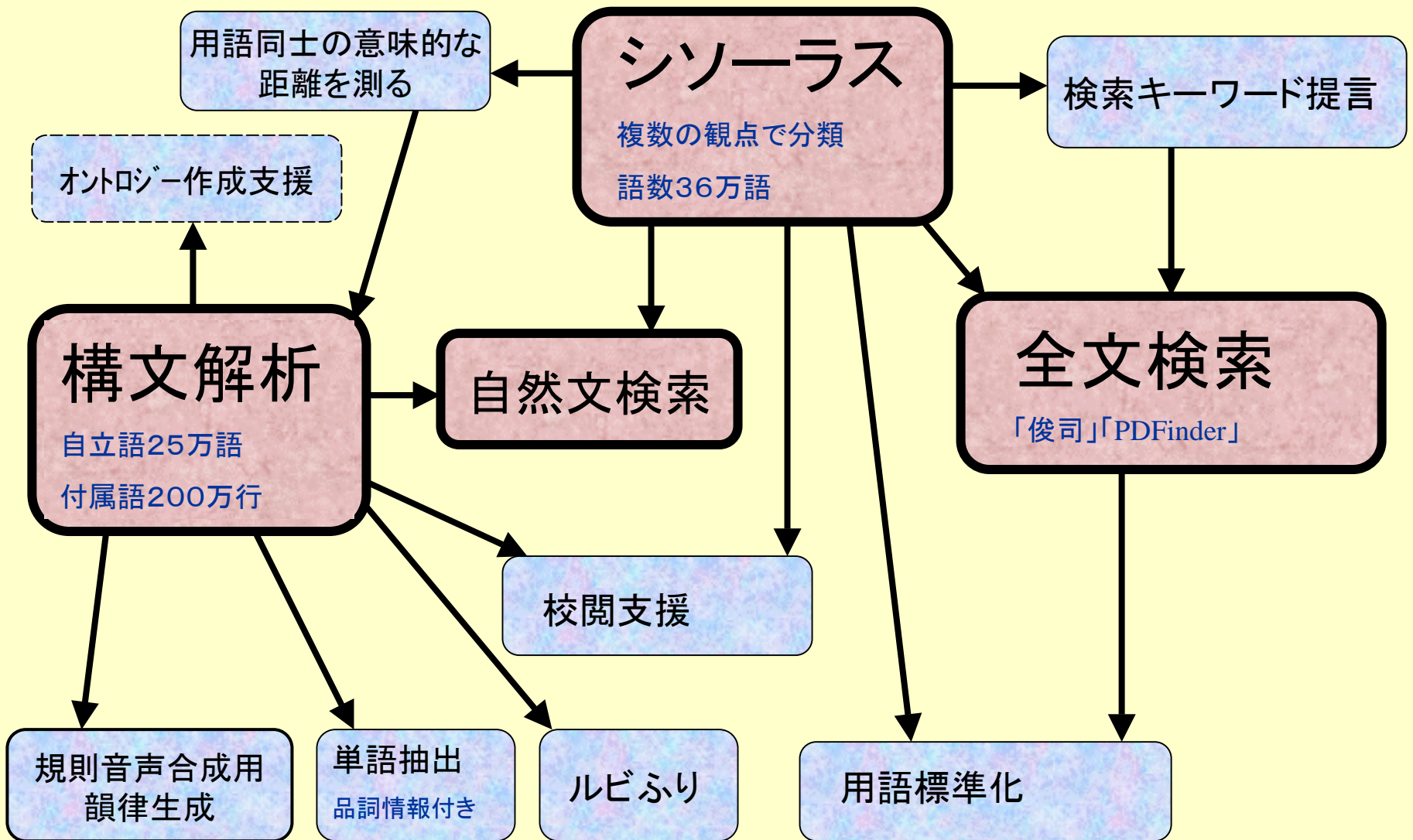
PDFファイル全文検索プラグイン

「PDFfinder」

校閲支援システム

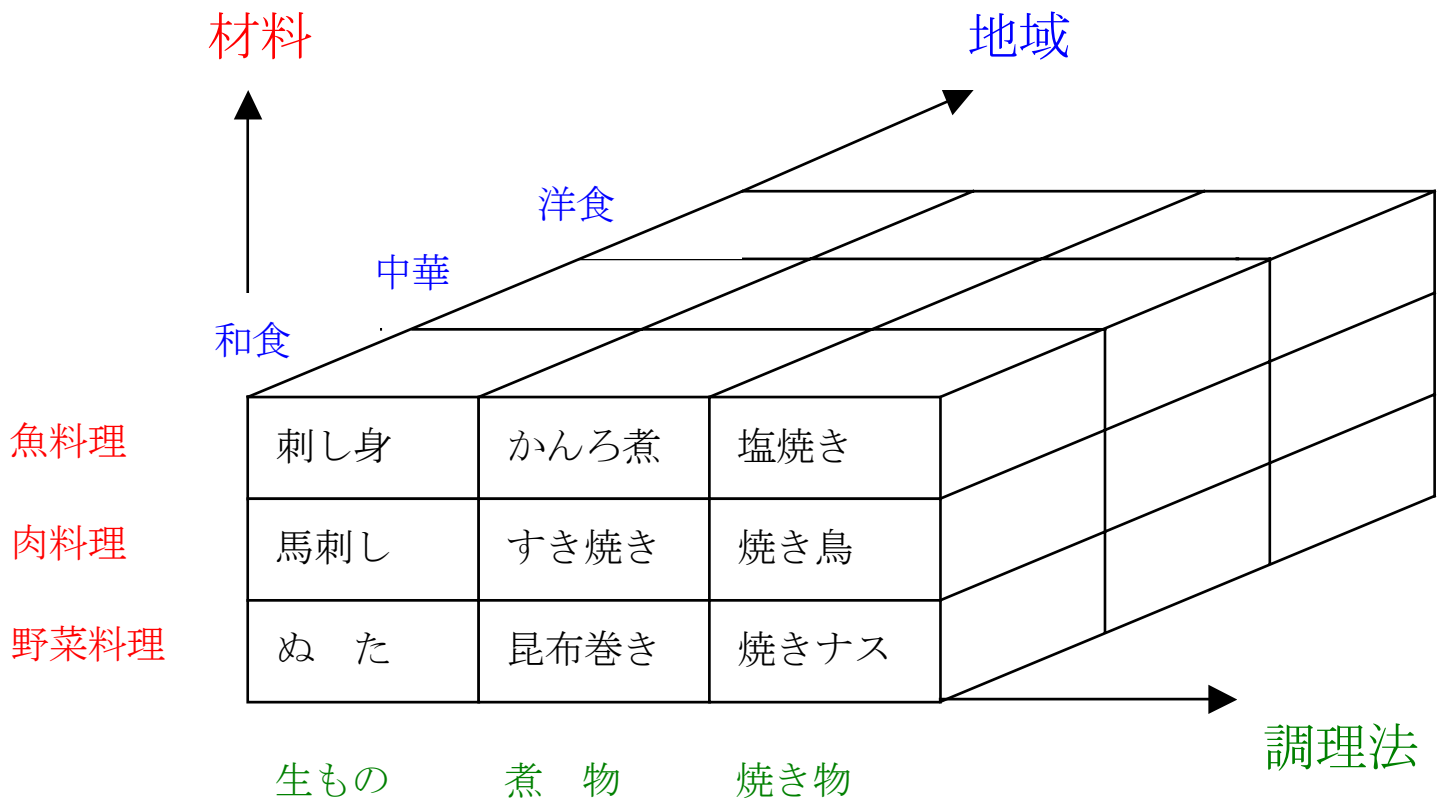
# 言語工学研究所のシステム体系

2007/11/13



# シソーラス

## 複数の観点で分類



# シソーラス(2)

異なり語数: 36万語

関係の種類

同義語、狭義語、広義語、関連語、反義語、語末一致  
差別語を出力しないようにしてある。

多義語は意味ごとに区別してある。

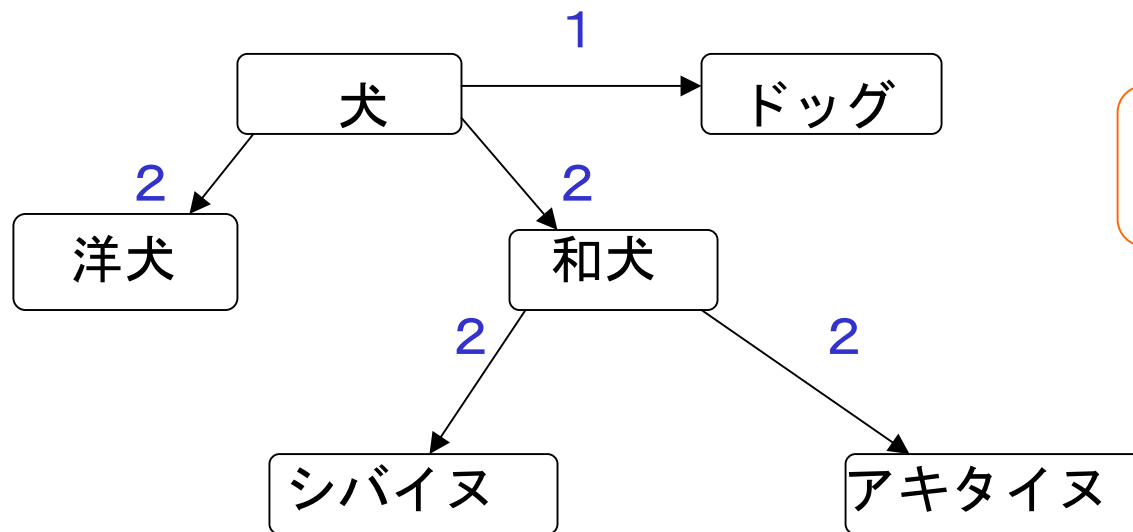
販売: パッケージ販売「デジタル類語辞典」

納入先: ライブドア、NHK、野村総研

有料サイト: <http://ruigo.jp/>

# 用語同士の意味的な距離

記事にある用語とキーワードとの距離による記事のスコアリング  
構文解析で係り先を決定する。



多義語は区別してある

3番目の用語との距離は2つの距離を加算する。

「犬」と「シバイヌ」の距離は「犬」・「和犬」の距離2に「和犬」と「シバイヌ」の距離2を加算して4とする。

## 意味的な距離(2)

構文解析で、係り先を決めるときに用いる。

用語同士の意味的な距離を測るシステムを利用して、検索結果のスコアリングが期待できる。

# 検索キーワード提言

検索に適切なキーを提言するシステムである。

例えば、

料理の記事には「料理」という言葉は出てこない。

「料理」の記事を探すためには

「レシピ」「下ごしらえ」「調味料」「食材」

「調理」「台所に立つ」・・・

# 全文検索

コンパクトでかつ高速。

パッケージ: PDF版 (PDFFinder)

最近JAVA版を出荷

納入先: 日経BP社

縮刷版

情報処理学会

創立45周年記念DVD

国立印刷局

地方新聞社

# 構文解析

- ・ 構文構造が読みやすいようまとめた形になっている。

解析についての  
問題は  
なかったということです。

解析に  
についての  
問題は  
なかったと  
いう  
ことです。

これを実現するために次のような項目数の辞書を用意している。

自立語辞書（上の例の赤の部分） 250,000語

付属語（上の例の青の部分） 3,000,000行

# 構文解析(2)

- ・なるべく正確な解析をするためにシソーラスによって用語と用語との距離を考慮して構造を決める。

クロールで—┘

泳いでいる—┘

彼女を—┘

見ていました。

望遠鏡で—┘

泳いでいる—┘ |

彼女を—┘ |

見ていました。

# 構文解析(3)

Eメールで取り交わされる文章が対象  
文章の意図を出力する。

良いか、悪いか。

否定、勧誘、疑問、依頼、希望...

シソーラスを用いて並列構造を決定している。

品詞分類が詳細である。

納入先:産総研、三菱総研、NHK技研

ニフティ(形態素解析)

# 自然文検索

1. 記事を構文解析して、結果の係り受けのデータをデータベース化しておく。
2. 検索文も構文解析して係り受けにして検索する。

テキストマイニング、知的クローラーなどの応用が  
考えられる

# 規則音声合成用韻律規則生成

アクセント、発音記号に変換

数詞、助数詞もほぼ完べき

意味によるアクセントの違いにも対処

上野（地名） 平板

（人名） 頭高

前の語の意味により読み方を変えている。

推薦・人（にん）

社会・人（じん）

鼻濁音、口濁音を区別している。

納入先：NHK技研

# 校閲支援

指摘された誤りをクリックすると正しい用語に置き換える。

誤った部分だけでなく、通常の利用語もシソーラスにより他の用語に置き換えができる。

置き換えた語の活用語尾も自動的に調整する。

食指をそそられた → 食指を動かされた

共起関係も調べる。

台風に会った → 遭った

販売：パッケージ。

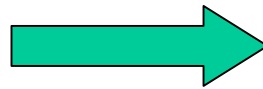
# 用語標準化(1)

統計などのため、標準の用語に変換

「計算機」

「電子計算機」

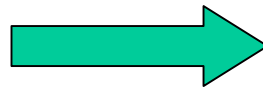
「パソコン」



「コンピューター」

「インタフェース」

「インターフェイス」



「インターフェース」

「的を得た」



「的を射た」

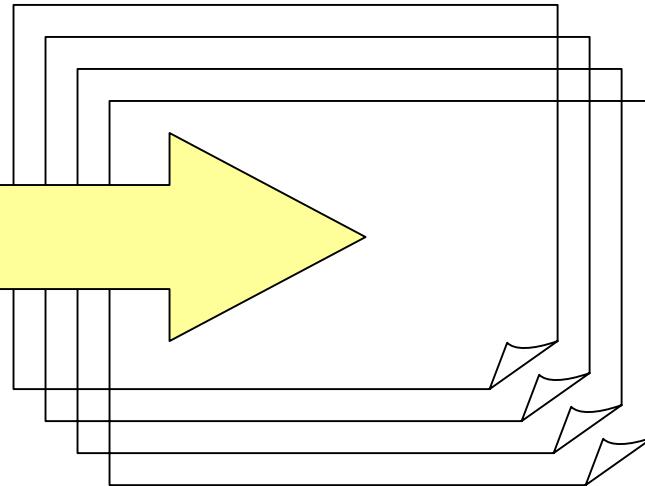
# 用語標準化(2)

異なった同義語が使われていないかを全文検索で  
複数の文書から検査

同義語辞書

米国  
USA  
アメリカ  
合衆国  
米

全文検索



文書1  
文書2  
文書3  
文書4  
.  
.

# 単語抽出

複合語の合成もしている。

世界的高学歴化 →

世界的、世界的高学歴化、高学歴化、学歴化

接尾辞で始まる語や、接頭辞で終わる語は出力しない。

× 世界的高、的高学歴、化

品詞情報付き

納入先：医学中央雑誌刊行会、農林水産省

# ルビ振り

個々の文字に対応した振り仮名をふる。

ご うち

碁打

小学校の学年に応じて部分的に振り仮名を振ることもできる。熟字訓にも対処している。

もの

果物      × くだ物      ○ くだもの

納入先: 朝日学生新聞社、凸版印刷

# 点字変換

点字に変換します。

東京へ

トーキョーエ

助数詞の連濁もサポート

1本

1ボン

2本

2ホン

3本

3ボン

鍼灸用語も登録済み。

マス空け付きで辞書に登録できる。

# オントロジー作成支援

## 自然言語処理向けのオントロジーを作成

### コーパスからボトムアップによる作成

1. コーパスを構文解析して、係り受けにして出力
2. 目的に応じた意味的依存関係を付加(学習機能付き)

共起辞書ファイルを用いる。

### トップダウンによる作成

日本語の正規表現辞書ファイルをもとに作成  
インスタンスをシソーラスから増強

# シソーラスのサイト

有料のサービスサイトを開設している。

アクセス数: 5000~1万件 / 1日

5回までは無料

<http://ruigo.jp>

# 会社概要

設立 1985年

資本金 1500万円

代表取締役 国分 芳宏

Eメール: [info@gengokk.co.jp](mailto:info@gengokk.co.jp)