

12. 意味・意図の解析

12.1 意味表現とは

12.2 規則による意味解析処理

12.3 統計的な意味解析処理

12.4 スマートフォンでの音声サービス

ニューラルネットワークによる意味解析

12.1 意味表現とは

- 意味表現の必要性

- 音声対話システムが応答・動作するためには、入力音声を機械可読な表現に変換する必要がある

- 意味表現の構成

- 発話タイプ

- ユーザの意図に対応
- マルチドメインシステムの場合は、ドメインを特定する情報も含む

- スロット情報

- 発話タイプに応じてスロットの組み合わせが決まる

これらを合わせて
意味フレームとよぶ

例)

```
[ask_weather,  
  location = 京都,  
  day = 明日]
```

12.1 意味表現とは

- 検索言語による意味表現と検索実行

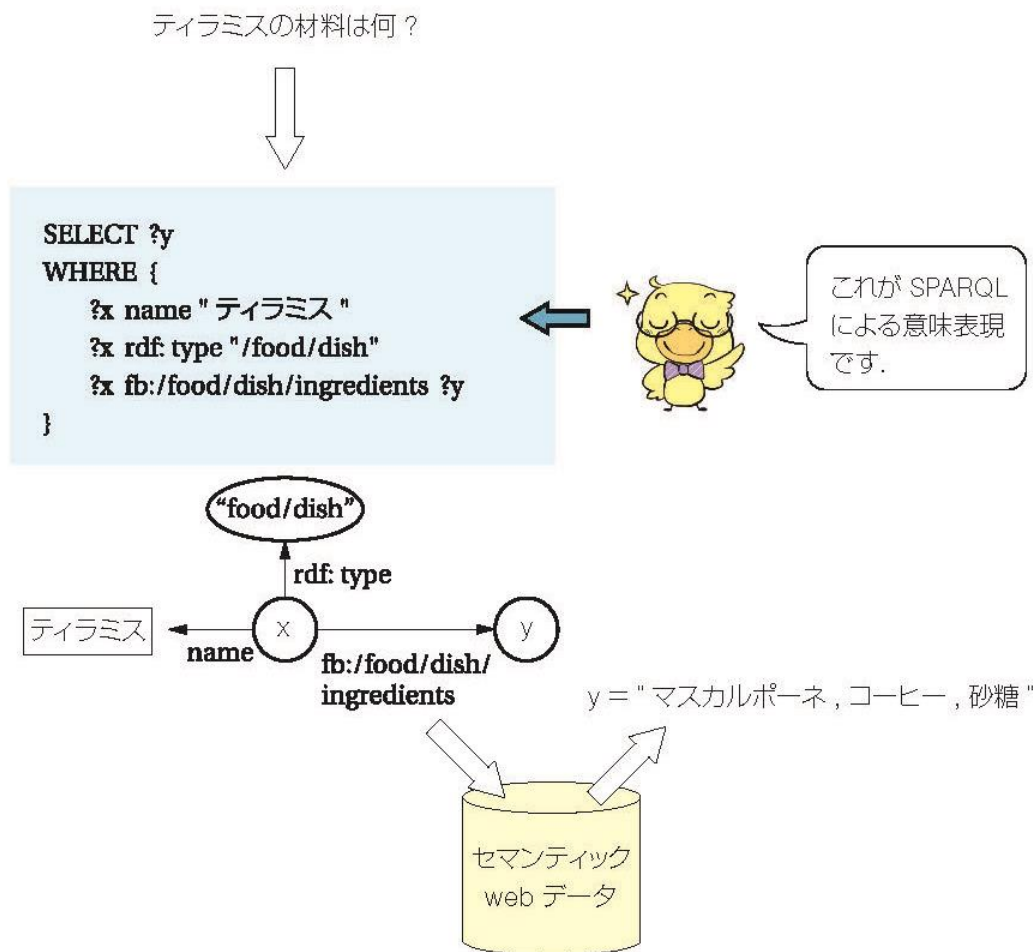


図 12.2 SPARQL を意味表現に用いた質問応答の例

12.2 規則による意味解析処理

- 規則による意味解析の適用範囲
 - 小語彙の音声対話システム
 - ユーザの入力発話が比較的定型的である場合
- 文法規則への意味表現生成規則の組み込み
 - 例) $\$文 \rightarrow \$表示 \mid \$設定 \mid \$検索$
 $\$検索 \rightarrow \$手段 \text{ で } 検索 \{[search, method=rules.latest()]\}$
 $\$手段 \rightarrow 住所 \{address\} \mid 名称 \{name\} \mid 履歴 \{history\}$
 - 構文木の葉から順に意味表現を組み上げてゆき、 $\$文$ に対応する意味表現が出力となる

12.2 規則による意味解析処理

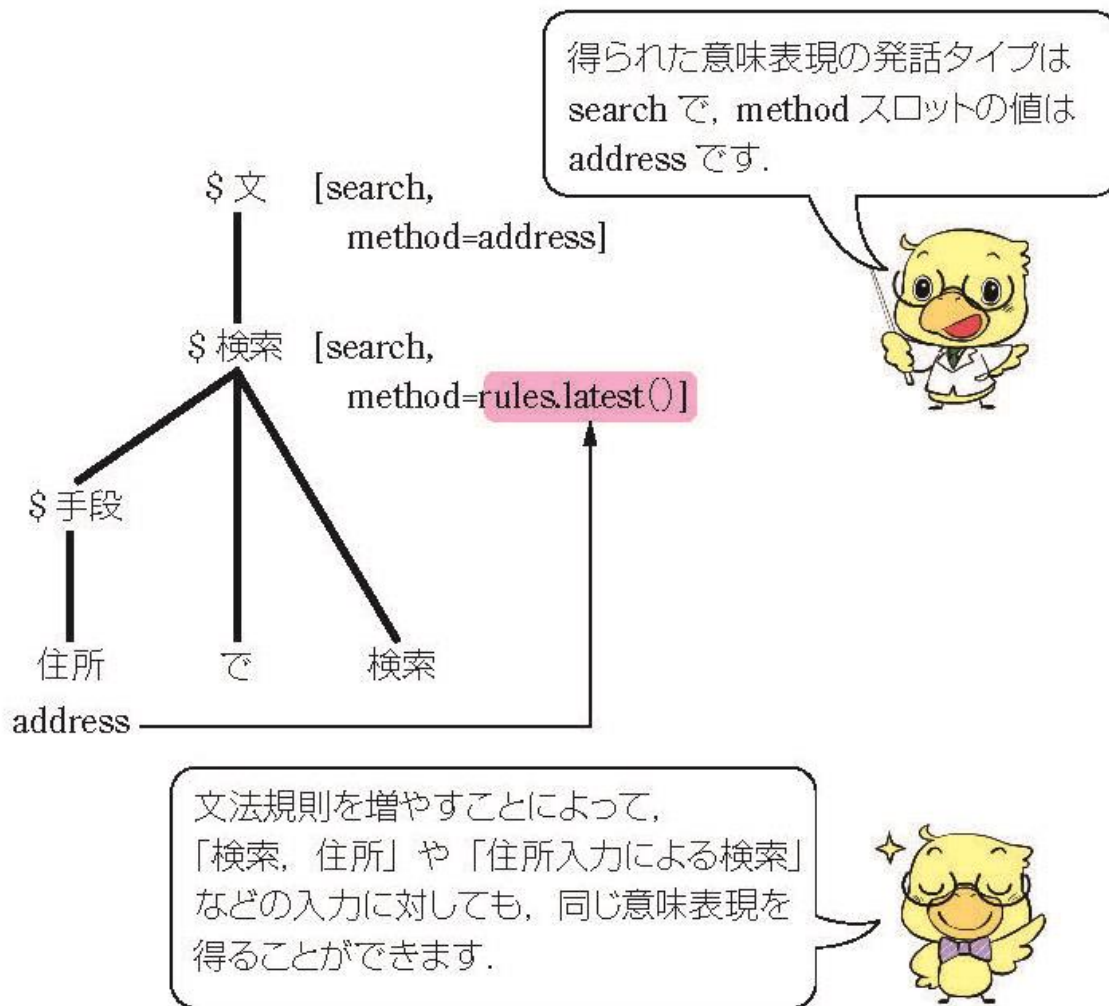


図 12.4 規則による意味解析の手順

12.3 統計的な意味解析処理

- 統計的な意味解析

$$\hat{M} = \arg \max_M P(M|W)$$

- 生成モデルによる解法

- 意味表現をフラット化（コンセプト列による表現）
 - 発話タイプとスロット名を出現順に並べる
 - 例) 「自宅に設定」 → $M=[location, set_direction]$

$$\hat{M} = \arg \max_M P(W|M)P(M)$$

コンセプトと単語
列の対応の確率

コンセプト列の
N-グラム

12.3 統計的な意味解析処理

- 識別モデルによる解法
 - 系列ラベリングによるコンセプトの抽出
 - 例) O B-Loc B-Loc I-Loc B-Loc I-Loc O O B-Tic B-Num O
えっと 京都 京都駅 から 東京駅 まで ください あ 自由席 2枚 です
 - 単語列から、事後確率最大のラベル列を求める

$$\hat{C} = \arg \max_C P(C|W)$$

$$P(C|W) = \frac{1}{Z} \prod_{i=1}^L \exp\left(\sum_{n=1}^N \lambda_n \phi_n(c_{i-1}, c_i, W)\right)$$

12.3 統計的な意味解析処理

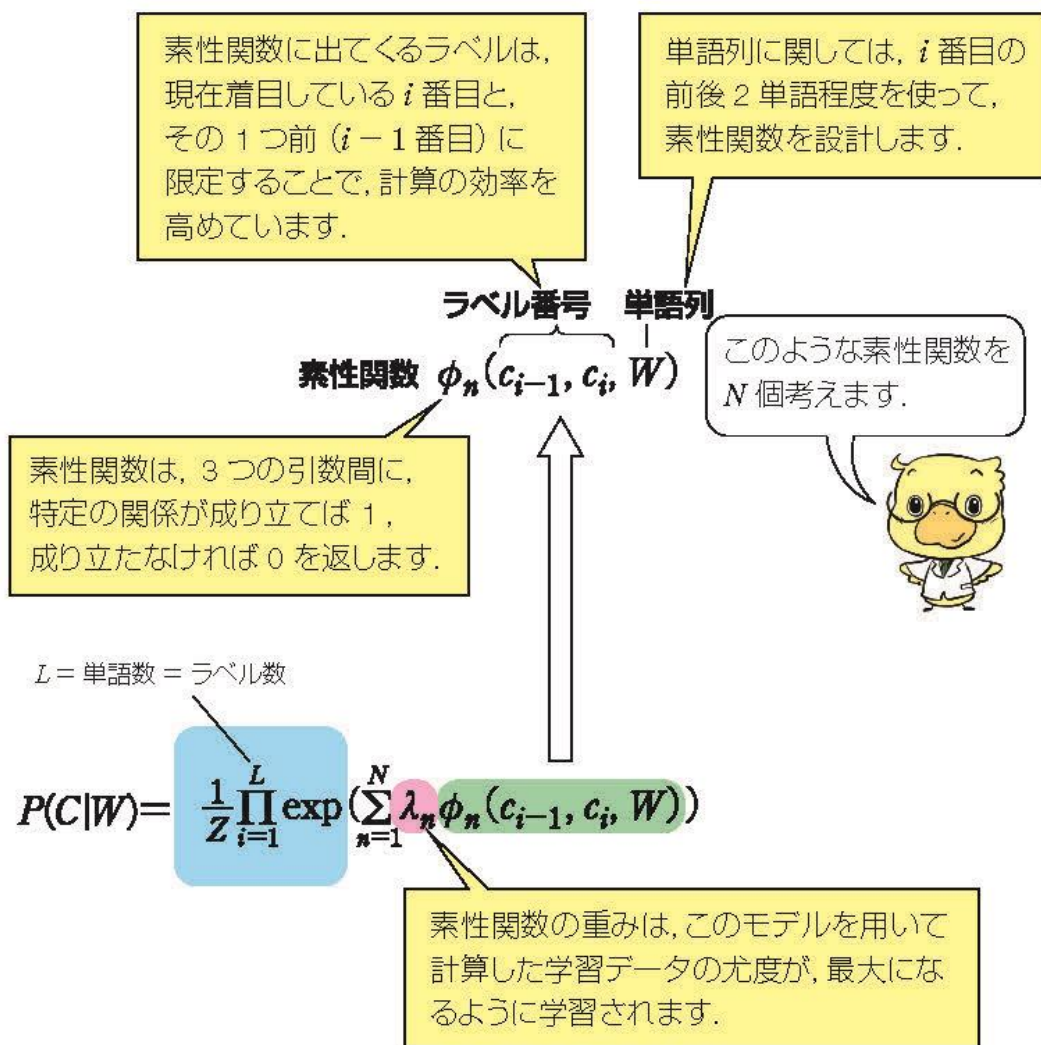


図 12.6 CRF の考え方

12.4 スマートフォンでの音声サービス

- スマートフォンでの音声インタフェースの実現例
 - 入力音声を端末側で特徴ベクトル系列に変換
 - 特徴ベクトル系列をクラウド上の音声認識サーバに送信し、認識結果を得る
 - 認識結果をクラウド上の意味解析サーバに送信し、意味表現を得る
 - アプリへのコマンドであれば、アプリを起動し、操作を実行
 - 検索等、専門エンジンへの入力であれば、各エンジンの対話インタフェースを呼び出す

12.4 スマートフォンでの音声サービス

- 「参照の揺れ」の問題
 - 検索対象に対して、略語・通称など、正式名称以外で参照された場合でも対象を特定する必要がある
 - 検索エンジンのログから確率モデルを作成する方法が有効

ニューラルネットワークによる意味解析

- Encoder-Decoderネットワークによる系列変換
[Kurata+ 2016]

