

# 知的情報検索実験システム IRIS

ICOT

第2研究室

(富士通株式会社)

秋山 幸司

## 位置づけ

応用分野を一つ定め、実際に動くシステムを試作することにより、自然言語理解技術の研究に寄与

- 対象分野を明確に意識した自然言語理解技術
  - ・『モデルに基づく理解』を大量の文に適用して評価する



数千件近いテキスト文の解析  
現実的な期間、労力でシステムを構築するための考察

## 背 景

### ○キーワード検索型テキストベースの普及

特徴： 文書情報(新聞, 文献抄録など)をそのまま格納  
内容を表す複数のキーワードを各テキストに付加

長所： データベース化(目的別データ抽出)よりも構築が容易

- ・情報抽出コストが小さい
- ・不特定な検索要求にも対応

問題： 高度なキーワード選定技能および対象分野の知識が必要

検索結果における不要情報の存在(結果の吟味が必要)

## 目 標

### — 欲しい内容を簡単に指定できる検索環境 —

### ○キーワード検索型テキストベースに対する知的検索

- ・日本語質問文による検索要求

キーワードの論理結合では表せない内容も自然言語に  
表現

- ・キーワードの選定の労力を軽減

特殊な検索技能や検索対象分野の知識を不要にする

- ・検索結果の吟味の労力を軽減

意味内容に基づく比較照合の試み



## アプローチ

### ○日本語質問文・テキスト文の解析

- ・対象分野のモデルに基づく係り受けや並列句の解析

### ○キーワード検索式の生成

- ・分野知識および検索技能を用いたキーワード検索式生成
- ・検索専門家の技能を持つエキスパートシステム

### ○検索結果の吟味

- ・質問文とテキスト文の意味構造の比較照合
- ・一致度に基づく検索結果テキストの順序付け

☆対象分野：情報産業界の新聞記事見出し文

## 知 識

○**分野モデル**：分野の実体の記述，実体間の関係（述語），  
制約条件

○**世界知識**：命題間関連知識，地理的知識，各種シソーラス，  
製品名・会社名の知識，など

○**単語辞書**：表記，文法属性，分野モデルとの対応関係

○**検索専門家の技能**：検索戦術，戦術適用条件

○**内容の比較照合の知識**：類似度規則，照合規則

# 構文意味解析部

○特 徴

- ・文節数最小法に基づく形態素解析
- ・意味的制約により，係り受け関係や並列句を，概念間関係に写像

○質問文解析部

- ・連体修飾節を高々一つ含む文を対象

○テキスト文解析部

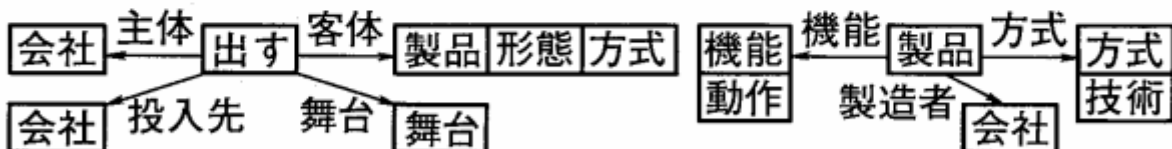
- ・見出し文の特徴に対処するヒューリスティクス
- ・暗黙情報(複数見出し間にまたがる関係)の解析

## 解析動作例

A社 , 16ビット パソコン 発売 ◆ ビジネス 市場 に 参入



分野モデル



## キーワード検索式生成部

### ○ 検索専門家の技能をルール化

- 検索戦術 : 検索処理を一步進めるための基本動作  
キーワードの展開や削除, 意味構造の変形
- 戦術適用条件 : 分野に応じた戦術適用の可否の判定  
適用の前後での類似性を示す距離値を設定

### ○ 意味構造間の類似度を示す指標値(距離)

- 質問文意味構造から一定の距離だけ離れた時, 展開・変形を終了

### ○ 距離指標値を評価関数とする動的な戦術適用順序決定

- 距離の近い展開を優先, 競合の解決(展開状態の分岐)

## 内容照合部

### ○ 基本的命題に注目した比較照合

- 『何が何をどうした』という命題/事象
- 述語間, およびその格の値である実体間, に分けた照合処理
- 関連度による推論制御, 順序付け

### ○ 実体間の照合処理

- キーワードレベル, 世界知識を用いた一致性判定

### ○ 述語間の照合処理

- 命題間関連知識を用いた前向き推論による類似述語への展開

## 評 価

### ○ 構文意味解析部

- ・ 質問文 : 解析成功率 90% (質問文70文)
- ・ テキスト文 : 解析成功率 70% (テキスト文2000件)

### ○ キーワード検索式生成部

- ・ 平均検索漏れ 32%, 平均不要情報 45%  
(質問文20文, テキスト文800件)
- ・ フリーターム方式キーワード検索の限界に近い性能

### ○ 内容照合部

- ・ 人間の順序付けと異なるテキストが, 検索結果に約 3 割出現

## 今後の課題

### ○ 知識獲得の自動化あるいは支援の必要性

- (現状)
- ・ 分野モデルの構築工数 2～4人月 (数百件の例文)
  - ・ 世界知識の抽出工数 1～3人月/千件

### ○ より完全な理解に向けて

- ・ 文脈解析, 談話理解, 類似性の探究

### ○ より高度で効率的な検索に向けて

- ・ 要約技術, より概念的なキーワードの自動付与技術