

7E-3

自然言語インタフェース・ツールの開発

—— 意味内容による文書の検索支援 ——

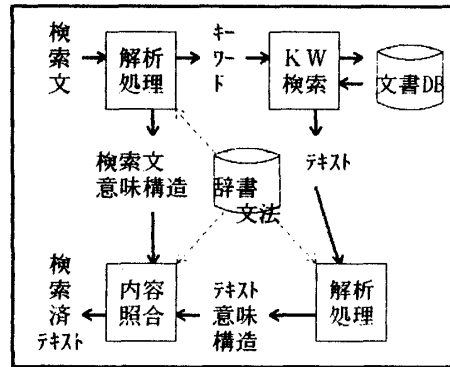
伊藤 博樹 西田 弘範 高橋 秋廣
富士通I7・71・ビ・研 朝東京放送

1. はじめに

我々は、自然言語インタフェースのツール作成を目指している。自然言語でのマシンとの会話で重要なのは、意味処理である。マシンで意味をどのように表現するか、文からどのように意味を抽出するかが鍵である。適用する分野及び用途を絞って、実用的な自然言語インタフェースを実現することにした。

そこで、応用分野をデータベースに限定し、さらに自然言語の利点が活かすため、文章データを対象として、それを検索文の意味内容により検索するシステムとし、加付例を試作した。

対象とする文章の内容は様々であるが、内容の分野毎に、また検索の用途毎にシステムを構築する。但し、システムの枠組みは共通化を図り、辞書・文法・限定分野知識等を外付けとした。



【図1 システム概念図】

2. システム概要

本作業では、東京放送所有のビワイブDBを対象とした。これは、ニュース用のビワイブを検索するためにその説明文をDB化したものである。

分野として報道・社会情報・スポーツ等があるが、加付例として比較的平易化しやすいスポーツを選んだ。

システム概念を図1に示す。検索文とテキストを意味解析して、共通の意味構造に変換し意味構造でマッチングすることにより意味内容による検索を実現する。その際、内容マッチングの処理時間短縮のため検索集合を検索文から抽出したキーワードで絞り込んでおいて、その部分集合テキストと内容マッチングする。

またテキストの意味構造への変換は、検索の度に処理せず、テキストの登録時に変換して、それを保存しておいて検索時に用いる。

文章の例(野球の場合)を図2に、検索文の例を図3に示す。

- (a)巨人対中日、3回戦、後楽園、0対2。
(b)巨人、槇原が先発、中日、小松が先発。
(c)4表落合が2塁打左へ、落合ホームへ向かうタッチアウト、中日1点先行。
(d)9裏、走者2塁で代打、原の飛球中堅、中日勝利。

【図2 一件文のテキスト例】

- ①後楽園で巨人の負けゲーム
②巨人が中日に負けた
③落合が槇原からヒット
④最終回の原の打撃

【図3 検索文例】

3. 実現方法

3.1 テキストの性質

対象とする文章には次のような特徴があった。

- ①読点句点の意味に用いられることがある。
②助詞や動詞の語尾などの省略が多い。
③暗黙の了解がある。

例えば野球の試合内容の記述の例では、
・「巨人対中日、0対2」で巨人が後攻であることと中日の勝ちを表す。
・「先発桑田。・・・3回落合の本塁打」で落合が桑田から打ったことを表す。

A design of the Natural Language Interface.
Hiroki ITOH Hironori NISHIDA
Fujitsu F.I.P. Corp.

Akihiro TAKAHASHI
Tokyo Broadcasting System, Inc.

### 3.2 日本語解析処理

解析処理は、単語分割・文節合成・構文解析・意味解析・簡易文脈解析から成る。意味解析で一文毎の解析を行い、簡易文脈解析で文間にまたがる暗黙情報を埋め込む処理を行う。

意味処理には、意味素性を用いた格文法による解析を採用した。図4に野球の分野での意味素性の例を示す。これは、野球という限定世界で用いられる単語（用語）を、その狭い世界での意味に注目して分類したものである。大きく名詞的概念と述語的概念に分け、全体で約120程度に分類した。次に、各述語的概念について、取りうる名詞的概念をその表層格マカとともに格フルムにまとめあげルルとして記述した。現在、野球に関するテスト500件を対象として、辞書3000語、文法500ルルの規模である。

### 3.3 意味表現

意味表現は、内容マッチングを考慮して、意味単位毎に平坦な構造とした。図2に示したテストの意味構造を図5に示す。図5に示すように、句点を区切りとした表層上の1文を、意味的なまとまりで複数のブロックに分けている。

### 4. おわりに

現在、解析可能な文章の表現としては、個人対個人の結果やチーム対チームの結果であるが、曖昧な表現（辛勝・惜敗など）や状況の表現（観客総立ち・うつつむく王監督など）は未対処であり、また他の分野（報道など）への適用も今後の課題である。

述語的概念	単語例	名詞的概念	単語例
投げる0	登板・投げる	主体	
1	好投・完封	チーム	巨人
2	降板	選手	原
打つ0	反撃・打つ	役割	
1	安打・本塁打	投手	先発
2	三振・凡退	打者	代打
守る0	守る	舞台	
1	ファインプレイ・さぼく	inning	回・表
2	失策・落球	アウト数	無死
結果0	分ける	走者	満塁
1	勝つ・ヨコヤ勝ち	打球方向	左前
2	負ける・落球	得点	2ラン

【図4 意味素性例】

【属性】	【値1】	【値2】
(a) CARD	巨人	中日
GAMECOUNT	3	
PLACE2	後楽園	
LOSE-TEAM	巨人	
WIN-TEAM	中日	
SCORE	0	2
(b) PITCHER	横原	
DEFE-TEAM	巨人	
P-ORDER	1	
(c) INNING	4	1ST-HALF
DEFE-TEAM	巨人	
OFFE-TEAM	中日	
BATTER	落合	
PITCHER	横原	
BATTING1	H-2	
DIRECTION	7	
INNING	4	1ST-HALF
HUMAN	落合	
RUNNINGO		
DIRECTION	2	
JUDGE2	TOUCHOUT	
TEAM	中日	
PROGRESSA	LEAD	
POINT	1	
(d) INNING	9	2ND-HALF
DEFE-TEAM	中日	
OFFE-TEAM	巨人	
BATTER	原	
PITCHER	小松	
A-ORDER	PH	
BATTING2	F	
DIRECTION	8	
RUNNER	2	
WIN-TEAM	中日	
LOSE-TEAM	巨人	

【図5 意味構造の例】