

# 言語処理技術はいかに使われているか?

Web情報アクセス技術専門委員会  
東京工業大学 大学院情報理工学研究科  
徳永健伸

## 委員会メンバー

新納浩幸  
(茨城大)

長谷川隆明  
(NTT)

池野篤司  
(沖電気)

奥西稔幸  
(シャープ)

小山誠  
(東芝)

河合英紀  
(NEC)

藤尾正和  
(日立)

齋藤孝広  
(富士通)

永井明人  
(三菱)

## 概要

- 2003年度
  - Webマイニング関連技術
  - Webサイトの検索・ナビゲーション機能
  - 米国における情報抽出技術の動向
- 2004年度
  - 言語処理技術の応用の実際
  - 言語処理の要素技術
  - 言語処理研究の動向

## 言語処理技術の応用：ヒアリング対象

- 生命情報学：小池麻子氏 (日立中研/東大)
- 特許情報検索：中田剛史氏 (特許庁)
- Blog検索 (1)：宮川達彦氏 (ライブドア)
- Blog検索 (2)：奥村学氏 (東工大)
- 医療文書管理：青江順一氏 (徳島大/ILU)

## 言語処理技術の応用：生命情報学

- 応用例
  - 専門用語データベースの構築
  - 相互作用と昨日情報の抽出
- 言語処理への期待
  - 精度が重要
  - 照応処理などの精度向上
  - データベース構築のノウハウ

## 言語処理技術の応用：特許情報検索

- 特許情報検索の入力：分類コード，キーワード
  - 分類コード
    - IPC, FI, Fターム → 人手による付与
  - キーワード
    - ゆれの問題→サーチ戦略ファイル
- 言語処理への期待
  - 分類コードの自動付与
  - 分類体系の自動構築
  - 概念検索

## 言語処理技術の応用：Blog検索

- 一般のWebページとBlogの違い
  - 情報の即時性 (更新頻度, RSS, ping)
  - 記述のスタイル
  - 個人の意見の比重が高い

## 言語処理技術の応用：医療情報管理

- 応用例
  - 音声による放射線科の読影入力
  - 感性表現解析
  - 同義語・表記のゆれの扱い
- 言語処理への期待
  - 辞書の構築・管理コストの削減



## 言語処理の要素技術：調査方法

- 調査対象：委員の所属企業が過去10年間にプレスリリースした製品・技術
- 分類観点
  - 言語処理の要素技術
  - 機能
  - 対象ドメイン
  - 年

## 要素技術 vs. 年 (1)

要素技術\年	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
形態素解析		1			1	2	2	2	5	3
構文解析						1			2	
文書構造解析			1					1		1
言語生成									1	
音声認識					2	3	6	5	3	3
音声合成					1	1	2	4	1	4
キーワード抽出					1			2	4	4
文書間類似度判定					1			2	3	2
重要文抽出							1			
情報検索	1	1	2	2	5	2	5	3	8	6
データマイニング							2		2	
クラスタリング							1	1	2	
情報抽出							2			1
情報推薦							2		1	

## 要素技術 vs. 年 (2)

要素技術\年	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
文書分類						2	4	1	1	
テキストマイニング						1	1	2	8	3
要約							1		1	
情報フィルタリング		1		1						
Webクローラ					1		1		1	
リンク解析					1		1			
機械翻訳		1			1	1	3	2	3	2
対話 I/F	1				1	1			1	1
知的ヘルプ					1					
手話合成						1				
手話翻訳						1				
文字認識			1		1	1	1	3	3	2
画像表示技術									1	1
情報可視化					1	2	1	1	4	1

## 対象ドメイン vs. 年

ドメイン\年	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
カーナビ						1	2	1		2
特許							1	1	2	
eコマース					1					
インターネット				1	3	3	5	1	1	
イントラネット	1		2				2	4	7	1
医療						1				
バイオ									1	2
CRM						1	3	1	3	3

## 機能 vs. 年

機能\年	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
翻訳		1			1	2	3	2	3	2
翻訳支援	1	1							2	1
コミュ・支援	1				1	2	4	2	2	2
アラート機能										1
知識管理			1		1		2		4	1
情報検索	1		1	2	5	1	6	3	5	4
情報提供		1							2	1
情報分析				1		1	2	2	9	3
入出力		2	1		1	5	4	6	3	5

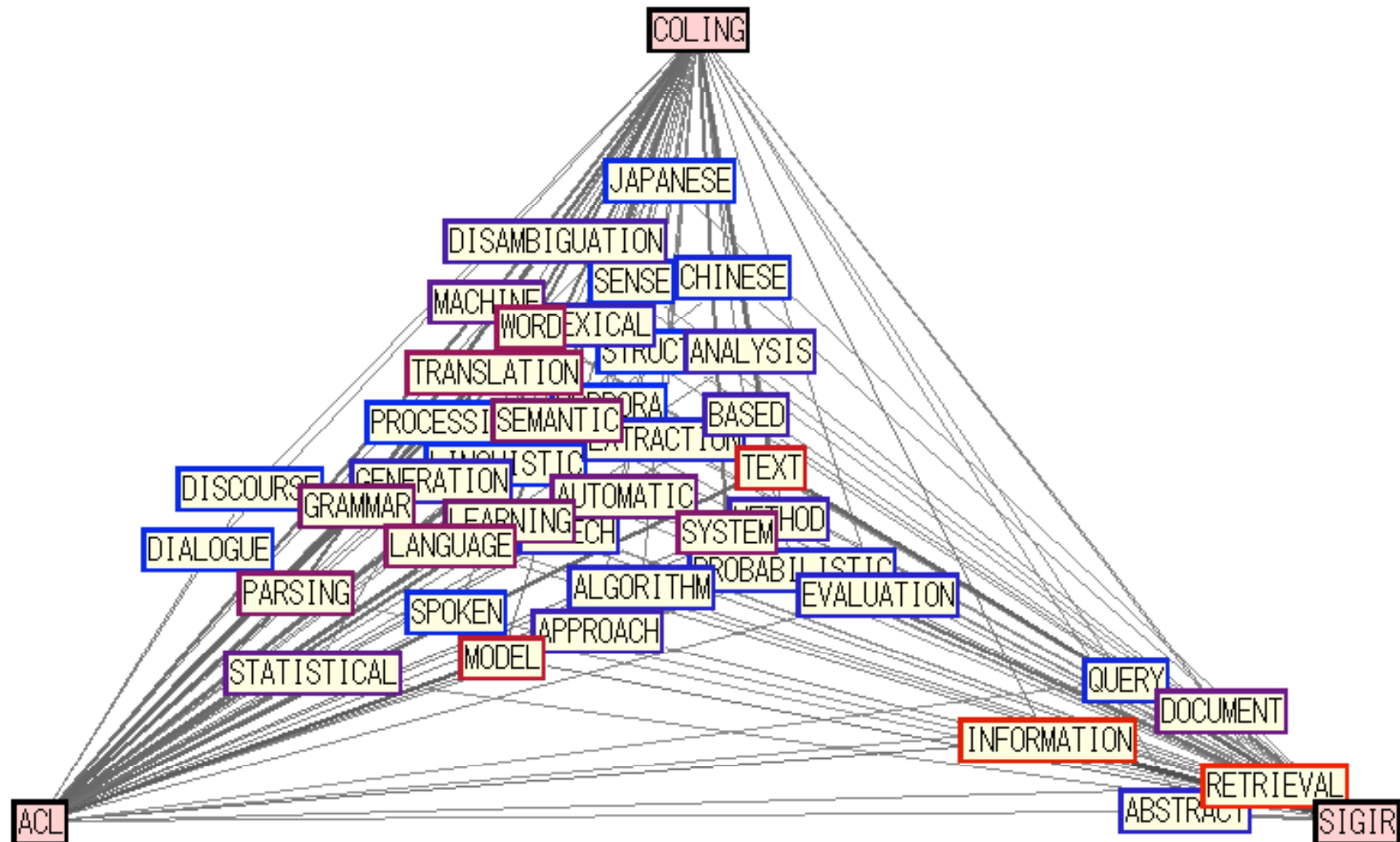
## 言語処理の要素技術：サマリ

- 1999, 2000年より情報検索・情報分析系のk/w  
が増加
- 検索系：単純検索→多言語, 情報分析
- 2000年：インターネット→イントラネット
- インターネット→情報検索,  
イントラネット→コミュニケーション支援
- 音声認識→広い応用  
文字認識→イントラネット

## 言語処理研究の動向：調査方法

- 1994～2004年の以下の国際会議の論文タイトルの分析
  - ACL Annual meeting
  - COLING
  - ACM SIGIR (2002年まで)
- 会議別，年別のタイトル中の語の出現傾向

# 会議別のキーワード





## 年別のキーワード (1994~1999)

順位	1994	1995	1996	1997	1998	1999
1	text	information	information	retrieval	retrieval	retrieval
2	grammar	retrieval	retrieval	information	text	information
3	system	model	language	model	information	system
4	retrieval	text	system	text	word	text
5	information	semantic	text	word	automatic	model
6	model	discourse	word	learning	language	speech
7	parsing	document	learning	semantic	model	parsing
8	lexical	evaluation	lexical	translation	learning	automatic
9	automatic	grammar	semantic	context	system	language
10	japanese	multiple	parsing	language	translation	semantic

## 年別のキーワード (1994~1999)

順位	2000	2001	2002	2003	2004
1	information	model	model	model	word
2	model	retrieval	text	word	machine
3	translation	information	language	translation	translation
4	text	text	information	text	model
5	language	language	retrieval	based	system
6	system	query	system	language	learning
7	word	system	translation	learning	semantic
8	parsing	web	document	semantic	text
9	retrieval	approach	word	spoken	information

## 研究論文の動向：要素技術 vs. 年 (1)

要素技術\年	94	95	96	97	98	99	00	01	02	03	04
形態素解析(71)	9		11	5	10	3	11		9	6	7
構文解析(179)	33	4	23	4	16	12	25	7	26	8	21
文書構造解析(1)					1						
言語生成(9)	1		1	1	1		2		2	1	
音声認識(11)			1		2	2	3	1	1		1
音声合成(3)	1				1						1
キーワード抽出(4)				1	2				1		
文書間類似度判定(0)											
重要文抽出(2)			1						1		
情報検索(367)	30	22	47	35	50	48	34	40	51	6	4
データマイニング(0)											
クラスタリング(39)	1	1	5	4	6	3	6	3	6	1	3
情報抽出(84)	8	1	7	3	6	2	14	3	16	9	15
情報推薦(2)									2		

## 研究論文の動向：要素技術 vs. 年 (2)

要素技術\年	94	95	96	97	98	99	00	01	02	03	04
文書分類(55)	5	4	5	4	5	3	6	6	13		4
テキストマイニング(14)					1	4	1		2	2	4
要約(47)		2			4	3	11	6	8	2	11
情報フィルタリング(29)	2		5	1	2	5	6	2	4	1	1
Webクローラ(1)								1			
リンク解析(4)								3	1		
機械翻訳(190)	20	3	21	9	22	8	28	9	29	14	27
対話 I/F (12)		1	1		1	1			2	4	2
知的ヘルプ(1)	1										
手話合成(0)											
手話翻訳(1)										1	
文字認識(7)	1		1	1	2		1		1		
画像表示技術(0)											
情報可視化(17)	1	1	5		2	2	3	1	2		

## 研究論文の動向：要素技術 vs. 年 (3)

要素技術\年	94	95	96	97	98	99	00	01	02	03	04
文法(136)	26	5	16	3	16	8	15	9	17	5	16
対話処理(125)	17	6	14	5	20	7	11	5	13	9	18
語義曖昧性解消(47)	3	2	5	3	4	2	3		8	3	14
質問応答(37)							6	6	13	3	9
照応解析(33)	4	2	5		5	1	2	1	4	2	7
単語分類(23)	1		7		3		3	2	3	1	3
語彙知識獲得(18)	2	1	2		2	1	3		2	2	3
単語アライメント(16)		1	2		1		2			1	9
固有表現抽出(15)					1		2	2	8	1	1
言い換え(14)	2		1	1		2		1	5		2
トピック分析(14)		1				1	1	4	6		1
テキスト分割(13)	2				5	2	1	3			
テキスト生成(10)	2	1	2		3		1		1		
翻字(9)				1	1		2		1		4
スペル訂正(9)	1		2		2		1		2		1

## まとめと展望

- 言語処理技術の応用分野
- 言語処理の要素技術
  - 製品・開発技術の観点
  - 研究論文の観点

## 予告：2005年度の活動

- テーマ：知識/情報管理
  - 個人の情報管理
  - イントラの情報管理
  - インターネットの情報管理
- 手法
  - ヒアリング
  - 製品調査
  - 研究論文の調査