

コールセンターDB 構築・検索と質問回答作成支援システム

鈴木 晃司、永井 秀利、中村 貞吾、野村 浩郷
九州工業大学 情報工学部
〒820-8502 飯塚市川津 680-4
nomura@ai.kyutech.ac.jp

あらまし パソコン技術サポートを行うメールコールセンターの自動化に関して、本研究では、既存の質問応答メールデータをカテゴリ化した質問応答データベースを構築し、さらにカテゴリごとにカテゴリ回答文を作成し、未知の問い合わせメールに対して文書ベクトルを用いて類似度を計算する質問カテゴリ判定システムによって属するカテゴリに判定され、そのカテゴリのカテゴリ回答文を用いて回答を作成し、作成した回答を最終査読者に提示する手法とシステムを提案する。提案手法の有効性を確認するため、実際のシステムを作成して実験を行った結果、145個の質問カテゴリを用いて、90%の割合で正解の質問カテゴリを上位3位以内に推定した。このシステムは、メールコールセンターで実務に供用できる状態にある。

キーワード メールコールセンター、質問応答データベース、文書ベクトル、類似度判定

Call Center DB Building/Searching and Their Application to a Support System for Composing the Answers to Questions

Kouji Suzuyama, Hidetoshi Nagai, Teigo Nakamura and Hirosato Nomura
Kyushu Institute of Technology
Iizuka 820-8502 Japan
nomura@ai.kyutech.ac.jp

Abstract This article discusses an extension of the support system we have developed by which it generates an answer to a question at a mail Call Center. The mail Call Center receives inquiries from computer users by e-mails and then it replies to them automatically. We have already developed such system that can be actually used at the mail Call Center. The mail Call Center generates the reply by finding out similar questions stored in the Data Base (DB) and then composes the answer by arranging them. Although it could make almost satisfying answers, the DB construction was not necessarily fully satisfied. By this reason, we reconstruct the DB so that the retrieve of similar questions could achieve more correctly and efficiently. In addition to this work, we improve the similarity calculation that evaluates the similarity between a new question and data stored in the Data Base. As a result, we could recognize that the performance has been actually improved and then the system became to produce more adequate answer.

Keyword mail Call Center, Q&A Data Base, document vector, similarity evaluation

1. はじめに

近年のパソコンの普及により、アフターサービスとしてのパソコン技術サポートの要望や利用が増大している。多くのパソコン技術サポートセンターでは、主に電話で技術サポートを行う従来型のコールセンターに加えて、E-mail での問い合わせを受け付けるメールコールセンターを多く設置している。そうしたメールコールセンターでの技術サポートは通常無料サポートであるため、その運営に係わる人件費などのコスト負担は膨大なものになっている。それゆえ、メールコールセンターの自動化が強く求められている。

メールによる技術サポートの場合、質問の内容は既に計算機可読な状態にあり、サポート結果としての回答作成処理はバッチ処理的に行うことが許容される。また、FAQのように、送られてくる質問には過去に出現した質問に類似したものが何度も送られてくるといった傾向があり、過去に作成した回答サンプルを有効に活用することが可能である。こうした点は、自然言語処理技術を十分に活かしている特徴と言える。

しかしながら、メールコールセンターを自動化しようとした場合、その精度はほぼ完全に 100%であることが求められる。現在の自然言語処理技術でそこまでの精度を得ることはほぼ不可能と言えるため、最終的な回答の作成には、人の手による査読が不可欠である。

そこで本研究では、より現実的なシステムとして、パソコンユーザからの質問に対してシステムが少数の回答候補を提示し、それらから選択、査読して回答を仕上げるような支援システムの構築を目指す。実際のメールコールセンターの業務においては、過去の何万件もの質問・応答サンプルからの適切な例の検索が作業全体の内でもかなりの比重を占めているため、査読のベースとなる回答案の候補を精度良く絞り込んで占めることができればメールコールセンターの運営コストの大幅な削減が実現できる。

我々は、これまでの研究で、実際のメールコールセンターでの約三年間の実務により収集した三万件以上の最終査読済みの「質問・応答」データの中から約一万件を分析し、回答内容の類似性に基づいて二層でカテゴリライズした。(実際の質問応答データは、すべて最下層のカテゴリのみに格納されている。)

以前の研究で、我々は、ある質問文への回答候補がどの質問応答カテゴリに属するかを判別する方法を提案し、145 個の質問応答カテゴリを用いて実験した結果、86%の割合で正解の質問応答カテゴリを上位 3 位以内に推定した。

本論文では、以前の研究で提案した手法に改良を加え、適切な回答候補が属する質問応答カテゴリの推定精度を向上させる手法を提案する。そしてその手法に基づく判別の指標となる値を質問応答カテゴリに与えて質問応答データベースを構築し、メールコールセ

ンターにおける回答作成支援システム ACCESS (Automated Call-Center Service System) を作成した。このシステムは、メールコールセンターで実際に実務に使用できる状態にある。

なお本論文では、著作権や知的所有権等の制約などにより、データなどの具体的な内容は概略や近似的例示のみを示すに留める。

2. 質問応答カテゴリ

2.1 質問応答カテゴリの作成

質問応答メールデータとは、パソコンユーザから送られてきた問い合わせメールとそれに対する最終査読済みの回答文 (= 応答文) とのペアのことである。質問応答メールデータの類似性は、問い合わせメール部分と応答文部分とのそれぞれの類似性に基づいて判断される。本研究の目的は、質問メールに対する回答文例を適切に見つけ出すことであるから、応答文部分の内容(文章表現)の類似性から同一の内容の回答であると見なし得るかどうかを重視して、カテゴリを作成する必要がある。使用機種名のように、各質問に依存した要素は、ワイルドカード化した上で比較を行い、同一性を判断する。

本研究では、質問応答カテゴリを二層のツリー構造によって関係付け、上位層をブランチカテゴリ、下位層をリーフカテゴリと呼ぶ。類似した質問応答メールデータを同一リーフカテゴリに分類し、さらに相関関係があるリーフカテゴリは同一ブランチカテゴリに分類する。リーフカテゴリはブランチカテゴリに属する場合もあるし、直接ルートカテゴリに属する場合もある。ルートカテゴリはブランチカテゴリの上位カテゴリであるが、ツリー構造のルートノードであるので、「層」とはみなさない。すなわち、全体を三層構造とは呼ばないことにしている。この概略を図示すると図 1 のようになる。

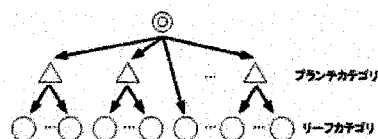


図 1 質問応答カテゴリの階層関係の概略図

これまでの研究において、我々は、10,135 件の質問応答メールデータを使用して質問応答カテゴリの作成を行った。この内、利用対象外データ 3,598 件を除いた計 6,537 件に対して、83 個のブランチカテゴリおよび 634 個のリーフカテゴリを得た。利用対象外データとしたものは、「メールコールセンターでのサポート対象外である」旨の応答などのように素直な形・内容のもではなかったものなどであり、再利用にはむかない、あるいは現在の研究の守備範囲を逸脱しているものなどで

ある。

2.2 回答文テンプレートの作成

カテゴリ構築の方針から、一つのリーフカテゴリに属する質問応答メールデータの応答部分は同じ内容であると見なし得るため、リーフカテゴリごとに共通回答文となる回答文テンプレートを作成する。

回答文テンプレートには、質問メール部分に記載された機種関連情報などに基づいて埋められるべきスロットが存在する。(もちろん、機種依存性などが全く存在しない応答の場合にはスロットは存在しない。) 査読用の回答例文は、システムがそれらを適宜埋めた上で提示する。

3. 質問文書特徴ベクトル

本研究では、サポート対象者であるユーザーから送られてきた問い合わせメールの質問文書の特徴を表すために、以下の属性を用いる。

- ・TF/IDF 値
- ・文言と用言との組での共起頻度
- ・文タイプごとの単語出現頻度

詳しくは文献[2]に譲るが、質問文書間の類似度を余弦類似度で求めるために、これらの属性値の集合を便宜的に多次元空間のベクトルとして扱い、「質問文書特徴ベクトル」と呼ぶこととする。

4. 質問応答カテゴリ特徴ベクトル

4.1 質問応答カテゴリの判別

問い合わせメールの質問文書がどの質問応答カテゴリに属するかを判別するために、本研究では質問応答カテゴリごとに「質問応答カテゴリ特徴ベクトル」を作成する。この質問応答カテゴリ特徴ベクトルと問い合わせメールの質問文書特徴ベクトルとの余弦類似度を求めて評価値とし、その値が最も大きいカテゴリに属するものと判断する。

4.2 従来の質問応答カテゴリ特徴ベクトルの問題点

従来の研究[2]では、質問応答カテゴリ特徴ベクトルとして、各カテゴリに属する質問文書の特徴ベクトルを単純に平均化したものを用いていた(図2を参照)。

この手法は簡便なものではあるため、学習および判別の計算コストを抑えることができるが、単純であるがゆえに、質問応答カテゴリ特徴ベクトルが必ずしも判別に最適なベクトルではないという問題が存在した。これは、各カテゴリの質問文書の特徴分布が必ずしも球状ではないという点に起因する。球状でないが故に、その平均点が分布中心からずれた位置となり、結果として判別誤りが増大することになる。

そこで本研究では、学習コストの増加を抑えつつ、より判別に敵した質問応答カテゴリ特徴ベクトルを作成する手法を提案する。

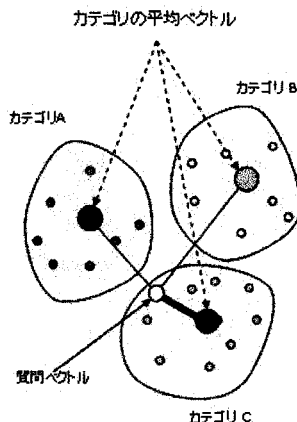


図2 従来手法での質問応答カテゴリの判別

5. 提案する質問応答カテゴリ特徴ベクトルの作成手法

5.1 初期質問応答カテゴリ特徴ベクトル

まず最初に、質問応答カテゴリ特徴ベクトルを得るための起点となるベクトルを求める。この初期質問応答カテゴリ特徴ベクトルには、従来手法における質問応答カテゴリ特徴ベクトルをそのまま用いるものとする。

5.2 特徴ベクトルの補正操作

設定されている質問応答カテゴリ特徴ベクトルがどの程度適切であるかは、質問応答カテゴリ構築の際に用いた質問応答メールデータによるクロスドテストで判定することができる。その際の誤判別傾向情報に基づいて質問応答カテゴリ特徴ベクトルを修正するのが「特徴ベクトルの補正操作」である。補正モデルの概念図を図3に示す。

本研究では、この誤判別について、質問応答カテゴリごとに次の2つのタイプを考える。

●誤判別A

対象の質問応答カテゴリに属しているはずの質問応答メールデータが、他の質問応答カテゴリへと判別されてしまうケース

●誤判別B

他の質問応答カテゴリに属しているはずの質問応答メールデータが、対象の質問応答カテゴリへと判別されてしまうケース

基本的には質問応答カテゴリ特徴ベクトルを、誤判別Aのグループには近づけ、誤判別Bのグループからは遠ざけるように操作すればよい。

ここで、対象の質問応答カテゴリの特徴ベクトル \vec{v}_c の補正式を、対象カテゴリに属する質問応答メールデ

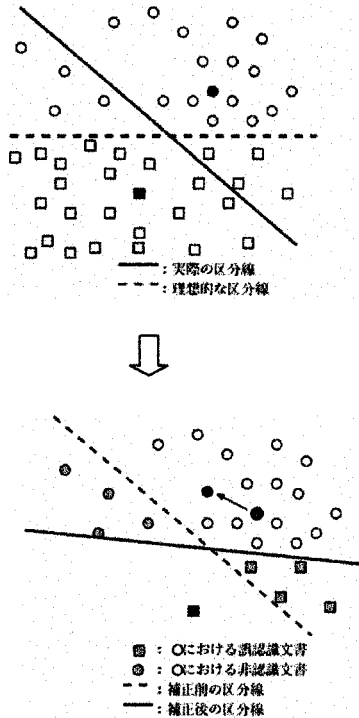


図3 質問応答カテゴリ特徴ベクトルの補正モデル

ータの総数 s 、誤判別Aとなった質問応答メールアドレスの総数 a 、誤判別Bとなった質問応答メールアドレスの総数 b 、誤判別Aとなった質問応答メールアドレスの平均ベクトル \vec{V}_a 、誤判別Aとなった質問応答メールアドレスの平均ベクトル \vec{V}_b を用いて次のように表す。

$$\vec{V}_c' = \vec{V}_c + \alpha \cdot \frac{a}{s+a+b} \cdot \frac{\vec{V}_c - \vec{V}_a}{\|\vec{V}_c - \vec{V}_a\|} - \beta \cdot \frac{b}{s+a+b} \cdot \frac{\vec{V}_c - \vec{V}_b}{\|\vec{V}_c - \vec{V}_b\|}$$

この式において、 $(\vec{V}_c - \vec{V}_a) / (\|\vec{V}_c - \vec{V}_a\|)$ と $(\vec{V}_c - \vec{V}_b) / (\|\vec{V}_c - \vec{V}_b\|)$ とが誤判別A、Bそれぞれのための補正方向の単位ベクトルであり、 $a/(s+a+b)$ と $b/(s+a+b)$ とがそれぞれの補正量制御である。誤判別数に応じた補正量とすることで、距離的に遠いサンプルや少数の例外的サンプルの影響を過大に受けてしまうことがないようにしている。

この式による操作を繰り返すことによって、理想的な質問応答カテゴリ特徴ベクトルへと漸近的に近づけることを目指すのが補正操作の目的である。正の数である α 、 β は、繰り返しの際の調整パラメータに相当する。

この値を大きくすれば調整速度は早まるが、理想に

十分に近づけきれなかったり補正結果の振動が生じたりする危険が高まることになる。本研究で作成したシステムにおいては、今回は α 、 β 共に値を 1.0 とした。

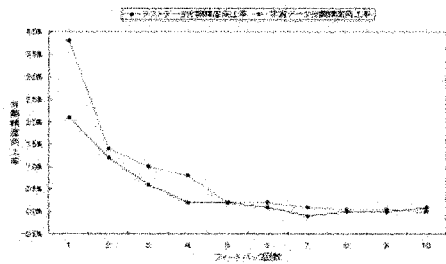


図4 ベクトル補正反復時の判別精度向上率の推移

図4に示すのは、実際に補正操作を反復した際の学習データとテストデータとのそれぞれでの判別精度向上率の推移である。学習データにおいて精度の向上が見られなくなったときにテストデータでも精度の向上が見られなくなるという傾向に基づき、本研究では学習データにおいて精度向上がみられなくなった時点で補正処理の反復を終了することとした。

5.2.1 特徴ベクトルの追加操作

質問応答カテゴリごとに一つの特徴ベクトルによって代表させるということは、カテゴリ境界が凹面を持たず、線型分離が可能であるという仮定を置いていることを意味する。しかしながら技術サポートにおける応答内容を考えた場合、同一の応答を返すべき状況であっても、その質問に至る際に顕現した現象は複数種類が存在し得る。そのような場合、同一カテゴリではあっても、質問応答メールアドレスの分布は複数種類の特徴傾向に基づく分布の重なりとなり、線型分離が不可能となる可能性が高い。

こうした問題に対処するために、何らかのクラスタリング処理を行って適当なクラスターの生成を目指すという方法も考えられるが、我々が必要としているのは精密なクラスターを作ることではない。適切な質問応答カテゴリを判別できさえすれば、クラスターとしては不十分であっても全く問題はない。

そこで本研究では、質問応答カテゴリ特徴ベクトルの補正操作によって解消し切れなかった誤判別の傾向に基づいて、新たな特徴ベクトルを追加する操作を行い、クラスタリング処理の代用とする。具体的には、誤判別A事例の平均ベクトル \vec{V}_a を新たな質問応答カテゴリ特徴ベクトルとして追加する。その概念図を図5に示す。

図6は、実際に追加操作を反復した際の学習データとテストデータとのそれぞれでの判別精度向上率の推移を示す。追加操作においても、補正操作の場合と同様に、学習データにおいて精度向上がみられなくなつ

た時点で追加処理の反復を終了することとした。

$$sim(\vec{t}, C) = \max_{\vec{v} \in S_c} \vec{t} \cdot \vec{v}$$

6. 回答作成支援システム

回答作成支援システムの回答例検索部のインターフェースを図7に示す。

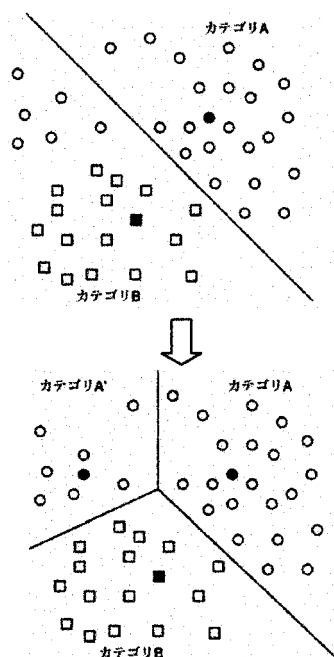


図5 質問応答カテゴリ特徴ベクトルの追加モデル

なお、特徴ベクトルの追加を行っても回答例作成のベースとなる質問応答カテゴリの数が増えるわけではないため、システム利用者の負担が増すことはない。

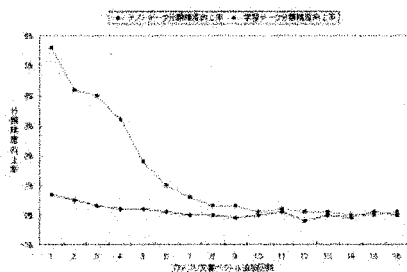


図6 カテゴリ文書ベクトル追加回数

5.2.2 特徴ベクトルを用いたカテゴリ判別

特徴ベクトルの追加処理を行った後は、同じ質問応答カテゴリに属する特徴ベクトルが複数存在することになる。したがってカテゴリの判別においては、カテゴリCに属する特徴ベクトルの集合を S_c として、次の式により得られた値、すなわち特徴ベクトルの内で最も高い類似度の値を評価値として用いる。

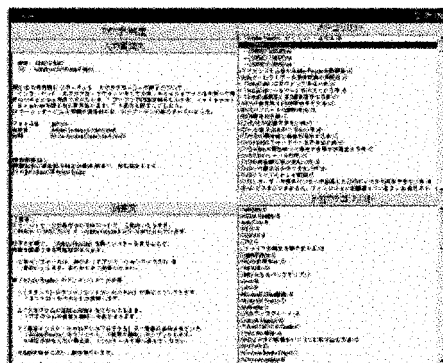


図7 システムの実行画面例

カテゴリ判別ボタンをクリックすると、質問文入力フォーム部に読み込まれた質問内容に基づいて、リーフカテゴリとブランチカテゴリのそれぞれについて評価値が高い順に表示される。表示された特定のカテゴリをクリックするとそのカテゴリに属している質問応答文（過去に最終査読まで行われ、応答として送られたもの）の一覧が表示され、その中から特定の質問応答文を選べば、別ウィンドウでそれを参照することができる。

7. カテゴリ判別実験

12件以上の質問応答メールデータが含まれるリーフカテゴリとそれらのブランチカテゴリを用いて質問応答カテゴリの判別精度の評価実験を行った。対象となったリーフカテゴリは145個、ブランチカテゴリは52個であり、含まれる質問応答メールデータの総数は4,023件である。

まず、すべてのメールデータを用いてクロスドテストを行った結果として、最上位および3位以内のカテゴリに正解が含まれていた割合を表1に示す。

表1 質問応答カテゴリ判別実験 (クロスドテスト)

	リーフカテゴリ		ブランチカテゴリ	
	一位に判定	三位以内	一位に判定	三位以内
従来手法	2,574 (68.0%)	3,496 (87.9%)	3,286 (84.7%)	3,866 (96.1%)
提案手法	3,705 (92.1%)	3,890 (96.7%)	3,850 (95.7%)	4,003 (99.5%)
改善率	75.30%	72.70%	71.90%	87.20%

従来の手法では 68%しかなかったリーフカテゴリでの一位正解率が提案手法では 92%まで上昇している。認識精度の改善率は 70%以上となっており、従来よりうまく分割平面を作成できているものと判断することができる。

次にオープンテストとして、メールデータを三分割し、内二つを学習データ、残りをテストデータとする実験を 3 回繰り返した。表 2 にテストデータに対する判別精度を示す。

表 2 質問応答カテゴリ判別実験(オープンテスト)

	リーフカテゴリ		ブランチカテゴリ	
	一位に判定	三位以内	一位に判定	三位以内
従来手法	2,505 (62.3%)	3,459 (86.0%)	3,202 (79.6%)	3,805 (94.6%)
提案手法	2,740 (68.1%)	3,589 (89.2%)	3,444 (85.6%)	3,922 (97.5%)
改善率	15.40%	22.90%	29.40%	53.70%

クローズドテストの場合ほどではないが、リーフカテゴリでの 1 位認識精度の改善率が 15.4%程度、他の場合も改善率 20%以上と、オープンテストの場合もかなりの改善が見られる。

この結果により、新しい問い合わせメールに対しても、ブランチカテゴリで上位 3 位程度まで調べれば、ほとんどの場合で最終査読者が最良と判定するような回答例を獲得できると期待できる。今回の実験結果では、オープンテストにおけるリーフカテゴリでの 1 位認識精度は 70%弱とまだまだ満足できる値とは言えなかったが、クローズドテストの結果を鑑みるに、学習に用いる質問応答メールデータを増やせば十分な精度を得ることができると思われる。

8. おわりに

本論文では、メールコールセンターでのパソコン技術サポートにおける応答メール作成作業の負担を軽減するために、適切な回答例候補を選出する精度の向上を目指した。

本論文で提案した手法を用いることにより、上位 3 位までの質問応答カテゴリ、すなわち 3 個の回答例候補を調べるだけで、90%弱の精度で適切な候補を特定でき、さらに、上位カテゴリのレベルで上位 3 位までに含まれる質問応答カテゴリを調べれば、97.5%の精度で適切な候補の獲得が期待できる。

したがって、本論文で提案した手法は十分に有効であり、そのままメールコールセンターでの実務に活用可能であると言える。

今回は手持ちの 3 万件強の質問応答データに対して約 1/3 のデータによりシステムを作成したが、クローズドテストとオープンテストとの結果の差異から、学習データ量の不足も確認された。したがって、手持ちのすべてのデータを用いてシステムを作成すると同時に、

学習データの増加に伴う判別精度の変化も調べる必要がある。

また、本論文の手法では、特徴ベクトル補正の際に誤判別集合が複数発生してしまった場合に、特徴ベクトル追加であまり適切とは言えない特徴ベクトルを追加してしまうケースが存在する。この点を改良し、より精度を高めることも今後の課題である。

本システムの名前は、ACCESS-mail である。この他に、現在、ACCESS-voice も開発中である。これについては、別の機会に述べる予定である。

参考文献

- [1] H. Nomura: "Mail/Voice Call Center Answering and Information Obtaining/Tidying", Proc. Of IPSI-2006 FRANCE, CD-ROM, 10 pages (2006)
- [2] 大元聡, 林暁紅, 永井秀利, 中村貞吾, 野村浩郷: "Q&A の自動回答支援システム", 情報処理学会研究報告 2006-NL-173, pp.47-52 (2006)
- [3] H. Nomura, M. Itoh, H. Nagai, T. Nakamura: "Information Extraction on Non-presupposed but Interesting Items in Addition to Presupposed Items", Proc. of IPSI-2005 SLOVENIA, CD-ROM, 6 pages (2005)
- [4] H. Nomura: "Information Retrieval and Integration of Relevant News Articles", Proc. of International Symposium on Machine Translation, NLP and Translation Support Systems, pp.4-9 (2004)
- [5] H. Nomura, H. Koga, H. Nagai, & T. Nakamura: "Text Summarization Based on Linguistic Function, Conceptual Relationships and Partial Contextual Constraints", Proc. of IPSI-2004 VENICE, CD-ROM, 6 pages (2004)
- [6] H. Nomura, S. Ohmoto, H. Nagai, T. Nakamura: "Information Extraction on Presupposed Items and/or Presupposed Topics from Text Documents", Proc. of IPSI-2004 STOCKHOLM, CD-ROM, 6 pages (2004)
- [7] H. Nomura, J. Akamatsu, H. Nagai, T. Nakamura: "Information Extraction on Presupposed Items from News Articles", Proc. of The Fifth International Conference on Electronic Measurement & Instruments, pp.626-630 (2001)
- [8] T. Nakamura, T. Matsubara, H. Nagai, H. Nomura: "A Method for Evaluating Uncertain Requirements in Travel Consultation Dialogues", Proc. of Third Natural Language Processing Pacific Rim Symposium, pp.290-295 (1995)
- [9] H. Nagai, T. Nakamura, & H. Nomura: "Skeleton Structure Acquisition of Japanese Law Sentences based on Linguistic Characteristics", Proc. of Third Natural Language Processing Pacific Rim Symposium, pp.143-148 (1995)
- [10] H. Nomura: "Information Extraction and Generation on Information Highway", Proc. of the 2nd Symposium on Natural Language Processing, pp.7-14 (1995)
- [11] S. Misumi, H. Nagai, T. Nakamura, H. Nomura: "Algorithmic Information Extraction from Natural Language Document", Proc. of Second Natural Language Processing Pacific Rim Symposium, pp.417-420 (1993)

(紙面都合により、参考文献は我々のもののみを掲載した)