

Hidden Markov Modelによる
専門用語の自動抽出

執筆者：甲田 英史

指導教官：新納 浩幸

平成12年3月1日

目次

第1章	序論	5
1.1	はじめに	5
1.2	処理手順	5
1.2.1	複合語抽出処理	5
1.2.2	専門用語判別処理	7
第2章	HMM	9
2.1	マルコフ過程	9
2.2	基本HMMの定式化	14
2.3	前向きパスアルゴリズム	18
2.4	Viterbi アルゴリズム	19
2.5	Viterbi アルゴリズムの例	21
第3章	本研究におけるHMMの設定	23
3.1	各系列における確率設定	24
3.1.1	状態遷移確率の割り当て	24
3.1.2	出力シンボルの条件付き確率	27
3.2	複合語判別処理	28
3.2.1	専門用語系列の遷移確率	28
3.2.2	一般語系列の遷移確率	30
3.2.3	その他の場合	32
第4章	抽出実験	33
4.1	実例による処理説明	34
4.2	詳細	40

第5章 考察	44
5.1 低正解率の原因	44
5.2 再現率について	45
第6章 おわりに	47
実験出力データ	50
プログラムリスト	104

目次

1.1	処理の流れ	6
2.1	2重マルコフモデルの例	12
2.2	ユニフィラーな有限状態情報源（決定性確率有限オートマトン）の例	13
2.3	簡単なHMMモデル	14
2.4	ナル遷移の除去	15
2.5	図2.3のHMMにおける $P(abb)$ のトリレス上の計算	19
2.6	図2.3のHMMにおける $P(abb)$ のViterbiアルゴリズムによる最適パス算出例	22
3.1	専門用語判別のためのHMMモデル	23
3.2	専門用語系列の確率割り当て	25
3.3	一般語系列の確率割り当て	26

表 目 次

4.1 実験結果	33
----------------	----

第1章 序論

1.1 はじめに

専門用語は、辞書作成・用語注釈・情報検索などの分野において有効利用されている。これより専門用語を収集整理することは重要である。また自動抽出が可能であれば、人手による煩雑な作業を軽減し、新たに増えつづける専門用語にも対応することができる。

従来はこうした自動抽出の方法として、専門用語の定義を行った後にその定義に基づいてその用語が専門用語かどうかの判別を行っていた。しかし実際には専門用語の定義は曖昧であり、結果の信頼性に欠けた [1, 2]。

本研究ではこの欠点を考慮して、機械学習手法を用いる。学習手法としてはHMM (Hidden Markov Model) を利用する。これにより状態遷移確率の比較によって専門用語と一般語との判別を行なうことができ、専門用語の定義を必要としない。また新たな専門用語が発生しても状態遷移確率は求められるので、比較判別が可能である。そこで必要となるデータは、コンピュータ用語辞典 [3] と電子版毎日新聞 [4] から抽出した。

1.2 処理手順

全体の流れとしては、まず文章を形態素解析し単語分割と品詞付けを行い、それを「複合語抽出プログラム」にかけ複合語を取り出す、その複合語をさらに「専門用語判別プログラム」にかけそれが専門用語であるかどうかの判別を行うという手順をとる (図 1.1 参照)。

1.2.1 複合語抽出処理

まず前提として抽出の対象とする専門用語は複合語とする。これは一語句のみで構成された専門用語の抽出は辞書との比較により既に可能なことと、新たに作られた専門用語は形態素解析を行うと、たとえそれが一語句であっても分割され複合語と分類されるからで

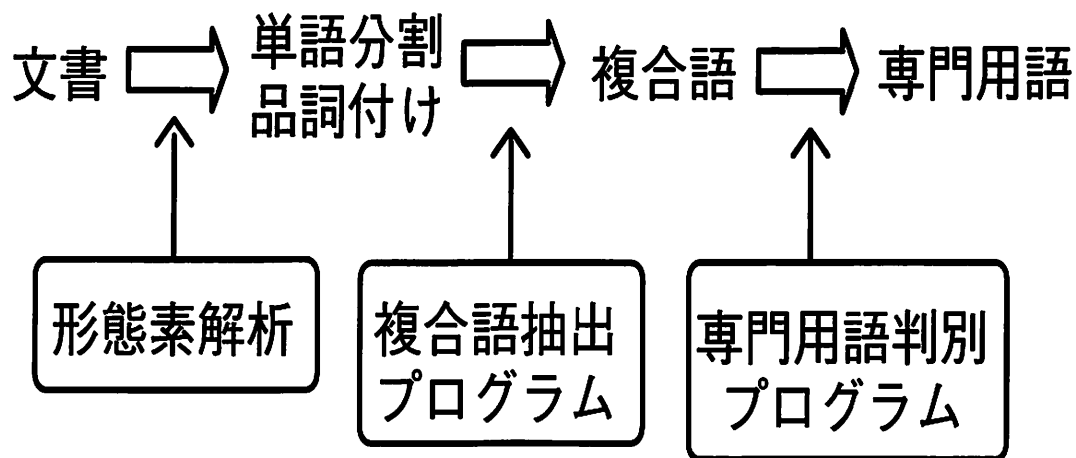


図 1.1: 処理の流れ

ある。

以下に各工程を記す。

1. 文章を形態素解析する

解析プログラム「茶釜」を用いて抽出対象の文書を形態素解析し、単語の分割と品詞付けを行う（この解析プログラムは既成）。

結果例：

エイズ	エイズ	名詞-一般	2*0*0
治療	チリョー	名詞-サ変接続	17*0*0
室	シツ	名詞-接尾-一般	28*0*0
が	ガ	助詞-格助詞-一般	58*0*0
最近	サイキン	名詞-副詞可能	16*0*0

2. 解析結果から複合語を抽出する

名詞の連続部分を探だし、それを今回作成した「複合語抽出プログラム」により複合語として抽出する。上記の例では「エイズ治療室」の部分である。同時に『18日』（数+助数詞）、『吉田氏』（人名+接尾）といった名詞の連続であっても明らかに抽出の

対象とならないものを、あらかじめ取り除く。

除外例：

18日----18(数),日(接尾-助数詞)

吉田氏----吉田(固有名詞-人名-姓)

氏(接尾-一般)

そのほかに代名詞(これ,それ),非自立(の,こと,もの),副詞可能(夜,すべて,いま)に属する名詞句も除外した。

逆に,外来語に多く見られる単語間に記号が入る語句やアルファベットでの略語を含んだ語句は複合語として抽出する。

名詞句以外を含む複合語：

システム・インテグレーション (記号・)

EAIツール (記号E, A, I)

クライアント/サーバ・システム (記号/,・)

1.2.2 専門用語判別処理

抽出した複合語が一般用語か専門用語かを判別することで専門用語の自動抽出が行える。この判別を機械学習手法を用いて行う。機械学習手法としては種々のクラス分類手法が利用できる [6]。ただしこの問題のタイプにあった手法を適用するべきである。本研究ではHMM法を利用する。

本研究のHMMではそのパラメータ推定のため多数の一般用語と専門用語が必要になる。今回は一般用語として電子版毎日新聞一年分,専門用語はコンピュータ用語辞典(CD-ROM) [3]から,複合語のみを抽出したものを使用する。

本研究では扱う専門分野を情報処理の分野に設定する。その専門用語はコンピュータ用語辞典から複合語となった見出し語だけを機械的に取り出して使用した。結果,110,543種類の専門用語を用意できた。また新聞記事にあらわれる複合語は一般用語と考え,毎日新

聞の94年度1年分の記事 [4] から前章で説明した複合語処理を経ることで498,682種類の一般用語を用意できた。

第2章 HMM

2.1 マルコフ過程

まず始めに、HMM法 (Hidden Markov Model, 隠れマルコフモデル) の基礎となるマルコフ過程について述べる。

時刻 t における確率変数を $Y_t (t = 1, 2, \dots)$ 、観測値を y_t としよう。 $Y_1 = y_1, Y_2 = y_2, \dots, Y_t = y_t$ と知ったときの、ナル確率が次式で与えられるとき、 n 重マルコフ過程という。

$$P(Y_{t+1} = y_{t+1} | \mathbf{Y}_1^t = \mathbf{y}_1^t) = P(Y_{t+1} = y_{t+1} | \mathbf{Y}_{t-n-1}^t = \mathbf{y}_{t-n+1}^t)$$

ここで、 $\mathbf{Y}_1^t(\mathbf{y}_1^t)$ は $Y_1, Y_2, \dots, Y_t (y_1, y_2, \dots, y_t)$ の時系列をあらわし、 $\mathbf{Y}_1^t = \mathbf{y}_1^t$ は、 $Y_1 = y_1, Y_2 = y_2, \dots, Y_t = y_t$ をあらわすものとする。特に、1重マルコフ過程のときには単純マルコフ過程という。

確率過程 $\{Y_t\}$ のとりうる値を有限個または可付番号無限個として、これを非負の数だけに限っても一般性を失わない。単純マルコフ過程においては、 $Y_1 = n_1, Y_2 = n_2, \dots, Y_t = n_t$ のとき、 $Y_{t+1} = n_{t+1}$ なる条件を考えれば、

$$P(Y_{t+1} = n_{t+1} | \mathbf{Y}_1^t = \mathbf{n}_1^t) = P(Y_{t+1} = n_{t+1} | Y_t = n_t)$$

が成立する。これが常に成立するものとして、

$$P_{ij}(t, t+1) = P(Y_{t+1} = j | Y_t = i)$$

と書くと、

$$P_{ij}(t, t+1) \geq \sum_j P_{ij}(t, t+1) = 1$$

である。 $P_{ij}(t, t+1)$ は、時刻 t に状態 i であったものが時刻 $t+1$ で状態 j に遷移する確率を表す。この遷移確率が t に無関係なとき、 Y_1, Y_2, \dots を定常な遷移確率をもつマルコフ過程という。以下、定常性を仮定して $P_{ij}(t, t+1)$ を単に P_{ij} と書くことにする。ま

た、初期状態確率 π を次式で定義する。

$$\pi_i = P(Y_1 = i), \sum_j \pi = 1$$

n 重マルコフ過程を考えると、一個の確率変数 Y_t のかわりに n 次元ベクトル $(Y_t, Y_{t+1}, \dots, Y_{t+n-1}) = \mathbf{Y}_t$ に着目すれば、新しいベクトル確率過程 $\{\mathbf{Y}_t\}$ が得られる。この過程では \mathbf{Y}_{t-1} の最初のベクトル成分は y_{t-1} であり、それ以外の成分は \mathbf{Y}_t に含まれている成分と同じであって、現在の事象はその一つ前の事象だけの影響下にあることとなり、確率変数の n 次元のベクトルを考えるだけで n 重マルコフ過程は単純マルコフ過程に帰着されることになる。

$\{\mathbf{Y}_t\}$ のとりうる値を L_1, L_2, \dots, L_S という S 個のシンボルとすれば、 $\mathbf{Y}_t = (Y_t, Y_{t+1}, \dots, Y_{t+n-1})$ のとりうる値はただか S^n (S^m の状態) のシンボルとなる。

定常な遷移確率 P_{ij} を持った単純マルコフ過程を考えよう。遷移確率 P_{ij} を要素とする行列

$$\mathbf{P} = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ P_{20} & P_{21} & P_{22} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

を遷移確率マトリクスという。一般に次の関係が成立する。

$$P_{ij}(m+n) = \sum_v P_{iv}(m) P_{vj}(n)$$

ここで、左辺はあるとき i であった場合に $(n+m)$ 回の遷移で j になる確率である。これを行列形式で書けば、

$$\mathbf{P}(m+n) = \mathbf{P}(m)\mathbf{P}(n) = (\mathbf{P})^{m+n}$$

となり、これは Kolmogorov-Chapman の方程式といわれている。 $t=1$ における初期状態の確率を \sum_i とすれば、 $Y_t = j$ となる確率は、

$$P_j(t) = \sum_i \pi_i \cdot P_{ij}(t-1)$$

となる。 $P_{ij}(t) > 0$ であるような t が存在することは、状態 i から状態 j へいつかは到達可能であることを意味し、任意の i, j について $P_{ij}(t) > 0$ ならこのマルコフ過程は正則であるという。どのような状態から出発しても何回かの後には必ずほかの任意の状態に移るこ

とのできるマルコフ過程はエルゴード的であるという。正則なマルコフ過程はエルゴード的である。正則なマルコフ過程では、初期上位の確率に関係なく、サンプル系列 $\{n_1, n_2, \dots, n_t\}$ に対して状態 j を通過する回数を $u_j(t)$ とすれば、状態 j を通過する確率 $\frac{u_j(t)}{t}$ は、 t を無限大にすれば一定値となる。これはエルゴードの定理と呼ばれている。

n 重マルコフ過程は状態遷移図で表現することができる。ベクトル $(y_t, y_{t+1}, \dots, y_{t+n-1})$ がとりうる値の一つを $(L_{i0}, L_{i1}, \dots, L_{in-1})$ とすればこれを 1 状態に対応させる。また、各状態について、その状態から 1 回の遷移で到達することができる状態（たかだか $(L_{i1}, L_{i2}, \dots, L_{in-1}, L_k)$, $k = 1, 2, \dots, S$ の S 個の状態）を方向を持った線で結ぶ。また、この遷移に対して、状態遷移確率を付随させ、さらに、一つのシンボル L_k を発生させるものとする。図 2.1 は $\{a, b\}$ の二つの出力シンボルをもつ単純マルコフ過程、2 重マルコフ過程の状態遷移図である。状態遷移確率は、このマルコフ過程によって生成されうる長い訓練用サンプル系列を用いることによって容易に求められる。例えば、図 2.1 で N_A を状態 A を通った回数、 $N_A(b)$ をシンボル b が状態 A から生成された回数とすれば、

$$P(b|a) = \frac{N(a, b)}{N(a)} = \frac{N_a(b)}{N_A}$$

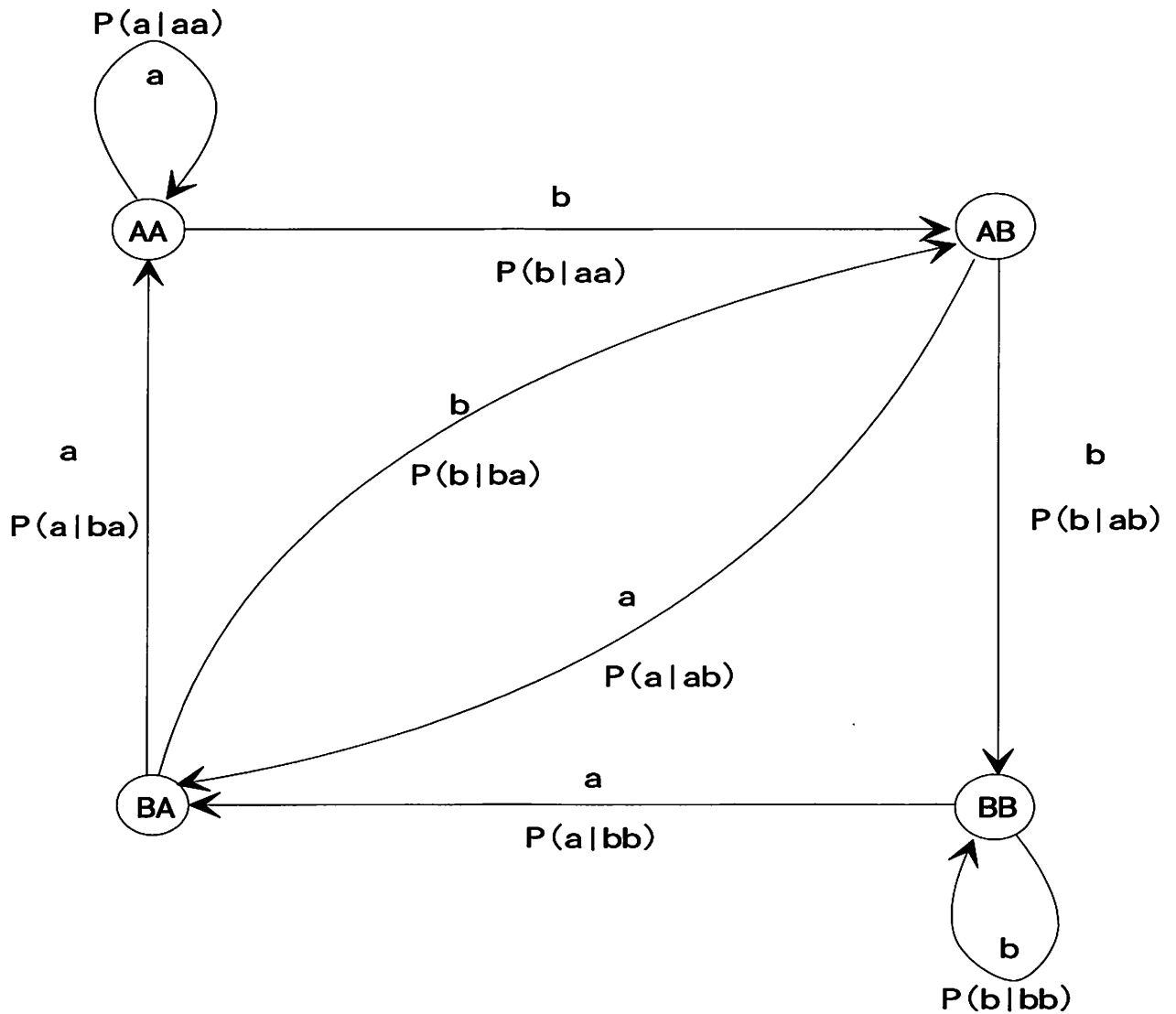


図 2.1: 2重マルコフモデルの例

同様に、図 2.1 で N_{AB} を状態 AB を通った回数、 $N_{AB}(b)$ をシンボル b が状態 AB から生成された回数とすると、

$$P(b|ab) = \frac{N(a, b, b)}{N(a, b)} = \frac{N_{AB}(b)}{N_{AB}}$$

となる。マルコフ過程のように、各状態から出ているすべての遷移がそれぞれ異なったシンボルを生成する有限状態情報源はユニフィラー (unifilar) と呼ばれる。ユニフィラーな有限情報源から生成されるシンボル系列を生成する確率は容易に求められる。たとえば、図 2.2 の例で、状態 1 が初期状態、状態 3 が最終状態であるとき、シンボル系列 $aabaaac$ の生

成確率は、

$$P(aabaaac) = 0.3 \times 0.3 \times 0.4 \times 0.2 \times 0.2 \times 0.2 \times 0.2 \times 0.1 = 2.88 \times 10^{-5}$$

となる。

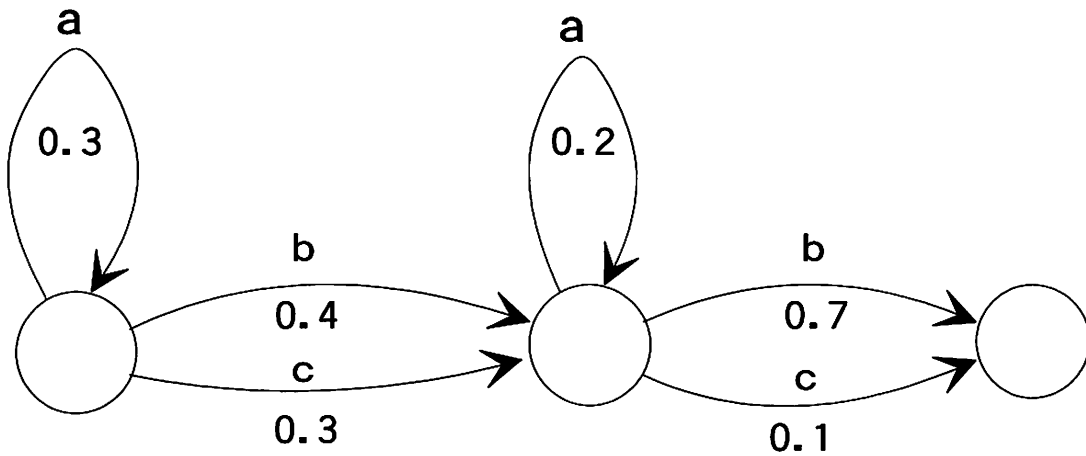


図 2.2: ユニフィラーな有限状態情報源 (決定性確率有限オートマトン) の例

2.2 基本HMMの定式化

HMM (Hidden Markov Model, 隠れマルコフモデル) は、出力シンボルによって一意に状態遷移が決まらないという意味での非決定有限状態オートマトンとして定義される (一般に、マルコフモデルは最終状態の概念はないが、今回の実験での単語分割に用いたマルコフモデルは初期状態、最終状態を設定する)。定義から当然HMMはユニフィラーではない。すなわち、出力シンボルが与えられても状態遷移系列は唯一に決まらない。観測できるのはシンボル系列だけであることから隠れ (hidden) マルコフモデルと呼ばれている。簡単な例を図 2.3 に示す。

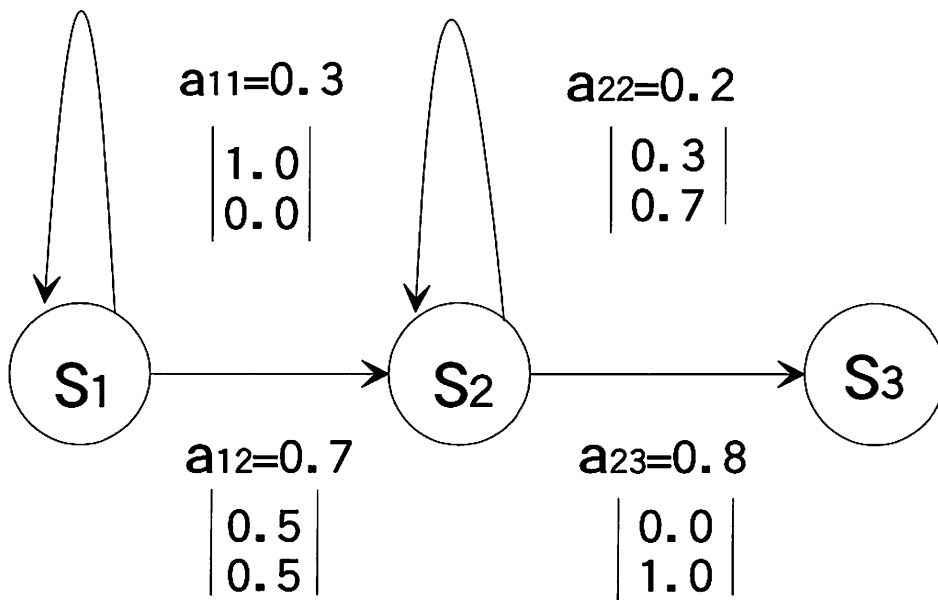


図 2.3: 簡単なHMMモデル

図中のアーク上のスカラ値は状態遷移確率を、アーク上のベクトル値はシンボル a, b の状態遷移による条件付きの出力確率を示している。すなわち、ベクトルの第一要素が a の、第二要素が b の出力確率を示している。この場合シンボル系列が abb であった場合可能性のある状態遷移系列は $S_1S_2S_3$ と $S_1S_2S_2S_3$ の二通りがある。今回は出力シンボル系列は最終状態に達するものとして扱う。通常HMMは全状態が最終状態となりうるが、複合語の単語分割では一部分の状態のみが最終状態となる。一般性を失うことなく、状態遷移にナル遷移 (シンボルを何も出力しない) は存在せず二つの状態間での遷移はたかだか一つ

と仮定できる. 前者の場合は、ナル遷移のときは空記号 λ が出力されたと考えればよい. また、図 2.4 で示すようにナル遷移を除去することもできる.

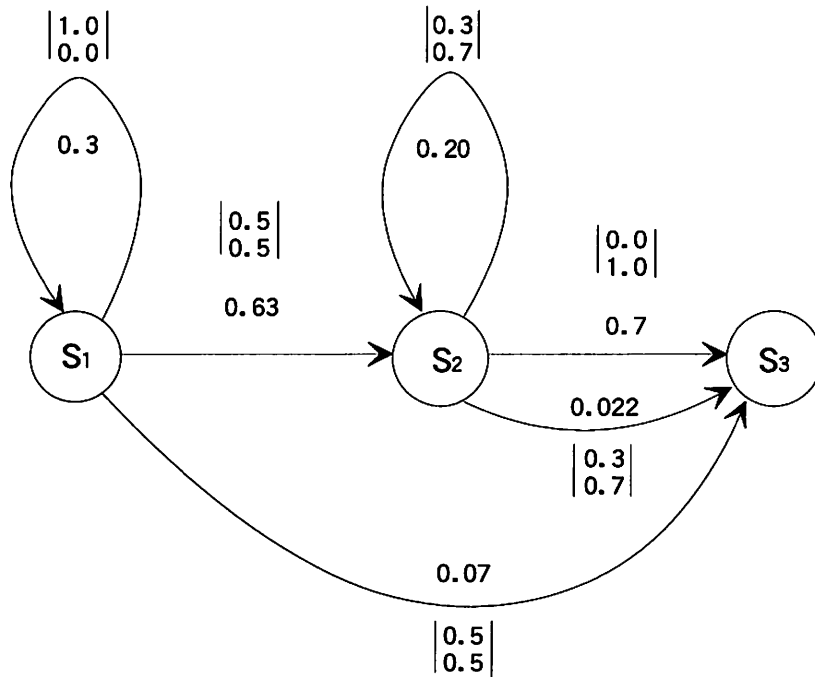
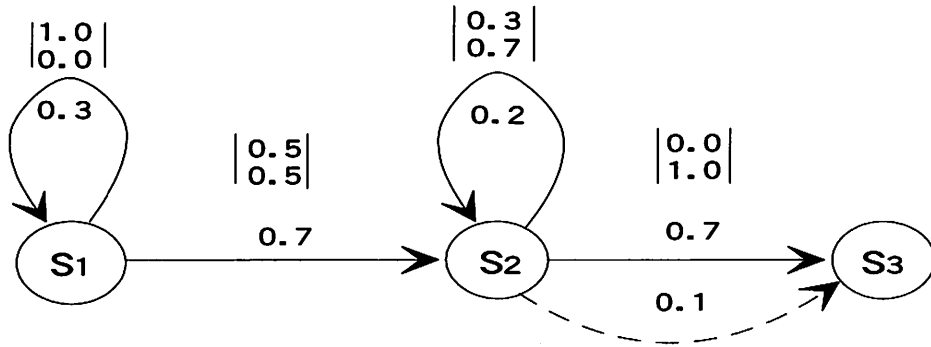


図 2.4: ナル遷移の除去

これによって後者が満たされなくなるが、二つの状態遷移 $S_2 \rightarrow S_3$ に対して、新たな状態 \hat{S}_3 を設けて、一方の遷移を $S_2 \rightarrow \hat{S}_3$ に変換し、状態 \hat{S}_3 の遷移先を状態 S_3 の遷移先に等しくすればよい. また、 y, x をそれぞれ出力と状態の確率変数値の系列としよう. 今考

えているHMMを \mathbf{M} とすれば、マルコフモデルの仮定から

$$P(\mathbf{x}_i|\mathbf{x}_i^{i-1}, \mathbf{M}) = P(\mathbf{x}_i|\mathbf{x}_{i-1}, \mathbf{M})$$

また、非決定性 (hidden) の仮定から、

$$P(\mathbf{y}_i|\mathbf{x}_1^i, \mathbf{y}_1^{i-1}, \mathbf{M}) = P(\mathbf{y}_i|\mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{M}) = P(\mathbf{y}_i|\mathbf{x}_{i-1}, \mathbf{M})$$

となる。式は状態遷移によって \mathbf{y}_i の出力確率が定まるとするもので、式は遷移元の状態によって定まるとするものである。混同のない限りHMMを示す \mathbf{M} の条件項を省くことにする。単語 w_i に対するHMMを上述の \mathbf{M} とすれば、式の \mathbf{Y}_{i-1}^i をあらためて \mathbf{y}_1^T とおけば、 $P(\mathbf{y}_1^T|\mathbf{M})$ は、

$$P(\mathbf{y}_1^T) = \sum_{\mathbf{x}} P(\mathbf{y}_1^T|\mathbf{x})P(\mathbf{x})$$

となる。また、

$$P(\mathbf{x}) = \prod_i P(\mathbf{x}_i|\mathbf{x}_1^{i-1}) = \prod_i P(\mathbf{x}_i|\mathbf{x}_{i-1})$$

から、

$$P(\mathbf{y}) = \sum_{\mathbf{x}} \prod_i P(\mathbf{x}_i|\mathbf{x}_{i-1}) \cdot P(\mathbf{y}_i|\mathbf{x}_{i-1}, \mathbf{x}_i)$$

となる。

図 2.3 の例で $P(abb)$ の値を式を用いて求めてみよう。

先にも述べたとおり $\prod_i P(\mathbf{y}_i|\mathbf{x}_{i-1}, \mathbf{x}_i)$ が 0 にならないのは二つの状態系列だけであったので、この各々の系列による値を $P_1(abb)$, $P_2(abb)$ とすれば、

$$P_1(abb) = 0.3 \times 1.0 \times 0.7 \times 0.5 \times 0.8 \times 1.0 = 0.084$$

$$P_2(abb) = 0.7 \times 0.5 \times 0.2 \times 0.7 \times 0.8 \times 1.0 = 0.0392$$

ゆえに、 $P(abb)$ は、

$$P(abb) = P_1(abb) + P_2(abb) = 0.084 + 0.0392 = 0.1232$$

となる。以上から、HMM \mathbf{M} は次の六つの組で定義される。

- \mathbf{S} : 状態の有限集合 ; $\mathbf{S} = \{s_i\}$
- \mathbf{Y} : 出力シンボルの集合

- \mathbf{A} : 状態遷移確率の集合 ; $\mathbf{A} = \{a_{ij}\}$; a_{ij} は状態 s_i から状態 s_j への遷移確率、ここで

$$\sum_j a_{ij} = 1.$$

- \mathbf{B} : 出力確率の集合 ; $\mathbf{B} = \{b_{ij}(k)\}$; $b_{ij}(k)$ は状態 s_i から状態 s_j への遷移の際にシンボル k を出力する確率. ここで、

$$\sum_k b_{ij}(k) = 1 (\text{離散 HMM})$$

$$\int_{-\infty}^{\infty} b_{ij}(k) dk = 1 (\text{連続 HMM})$$

- $\boldsymbol{\pi}$: 初期状態確率の集合 ; $\boldsymbol{\pi} = \{\pi_i\}$; π_i は初期状態が s_i である確率.

$$\sum_j \pi_j = 1$$

- \mathbf{F} : 最終状態の集合

ここで注意を要するのは、出力 \mathbf{Y} が有限集合の場合と無限集合の場合とがあることである。各々の場合について \mathbf{B} の表現方法が異なってくる。前者の場合は、出力確率分布は離散的でノンパラメトリックである。シンボルは有限集合だからあらかじめ \mathbf{B} を求めてテーブル化しておく。一方、後者の場合のそれは、ガウス分布のような(連続分布)で与えられる。連続分布の時は、シンボルは無限集合となり、あらかじめテーブル化しておけないので、出力シンボルが観測されるごとに $b_{ij}(k)$ を求めなければならない。

また y_i は時間的に等間隔に観測される場合と可変時間間隔ごとに観測される場合とがある。

\mathbf{Y} の出力形式以外に HMM に関して重要な基本問題として次の四つを挙げることができる。

1. モデルの設計 : 状態数や遷移先の種類、状態継続時間制御などどのような HMM を用いるかを決定すること。
2. モデルの評価 : モデル \mathbf{M} がシンボル系列 $\mathbf{y} = y_1, y_2, \dots, y_T$ を出力する確率(尤度) $P(\mathbf{y}|\mathbf{M})$ を求めること(認識時には、各モデルに対して、 $P(\mathbf{y}|\mathbf{M})$ が最大になる \mathbf{M} を決定する。

3. 最適状態系列の推定：モデル M がシンボル系列 y を出力するときの最も可能性の高い状態遷移系列を推定し、その系列に対する確率を求めること。
4. モデルの推定：訓練用シンボル系列 y を与えて、 $P(y|M)$ が最大になるようにモデル M のパラメータを推定すること。

本研究ではこのうち「最適状態系列の推定」を利用する。専門用語と一般語の2系列のモデルを用意し、各系列の状態遷移確率を比較することで、判別をおこなう。

2.3 前向きパスアルゴリズム

出力シンボル系列を与えた時、通ってきた状態遷移系列を推定することで単語分割が行なえる。推定する方法として全解探索がある。しかしこの方法では出力シンボルの数を n とすると組み合わせの数が 2^{n-1} となり、長い文字では推定のコストが高い。そこで、本実験では動的計画法の一種である Viterbi アルゴリズムを使うことにする。ここでは、Viterbi アルゴリズムの主要なモジュールである、前向きパスアルゴリズムを示す。

図 2.3 について考えてみる。入力が abb の場合、可能な二つの状態遷移系列にたいして式で示したように各々 5 回の乗算が必要であった。また、 $P(abb)$ の計算回数は 10 回の乗算と 1 回の加算であった。この計算は次のように書き換えられる。

$$P(abb) = [P(ab, S_1 \rightarrow S_1 \rightarrow S_2) + P(ab, S_1 \rightarrow S_2 \rightarrow S_2)] \times P(b, S_2 \rightarrow S_3)$$

このようにすると、全計算回数は 8 回の乗算と 1 回の加算ですむ。一般的に次に示す前向きパスアルゴリズムによって効率よく $P(y_1, y_2, \dots, y_T)$ を求めることができる。ここで $\alpha(i, t)$ を y_1, y_2, \dots, y_t を生成して状態 i に達する確率としよう。また、 F を最終状態の集合とする。すると、

$$P(y_1, y_2, \dots, y_T) = \sum_{i \in F} \alpha(i, T)$$

となる。 $\alpha(i, T)$ の算出アルゴリズムを示す。

前向きパスアルゴリズム

前向きパスアルゴリズムとは以下のようにして $P(y_1, y_2, \dots, y_t)$ を求めるアルゴリズムである。ただし、 $\alpha(i, t)$ は y_1, y_2, \dots, y_t を生成して状態 i に達する確率とする。

1. 初期化 すべての状態 i に対して $\alpha(i, 0) = \pi$
2. $t = 1, 2, \dots, T$ に対して 3, 4 を実行
3. すべての状態 i に対して 4 を実行
4. $\alpha(i, t) = \sum_j \alpha(j, t-1) a_{ij} b_{ji}(y_t)$
5. $P(y_1, y_2, \dots, y_T) = \sum_{i \in F} \alpha(i, T)$

これは、出力シンボル系列に対応する時間経過を横軸に、各状態を縦に並べて許される状態遷移を示すトレリス (trellis) の上で考えると理解しやすい。図 2.5 は図 2.3 の HMM で出力シンボル系列が abb の場合のトレリス上での $\alpha(i, t)$ の値を示している。

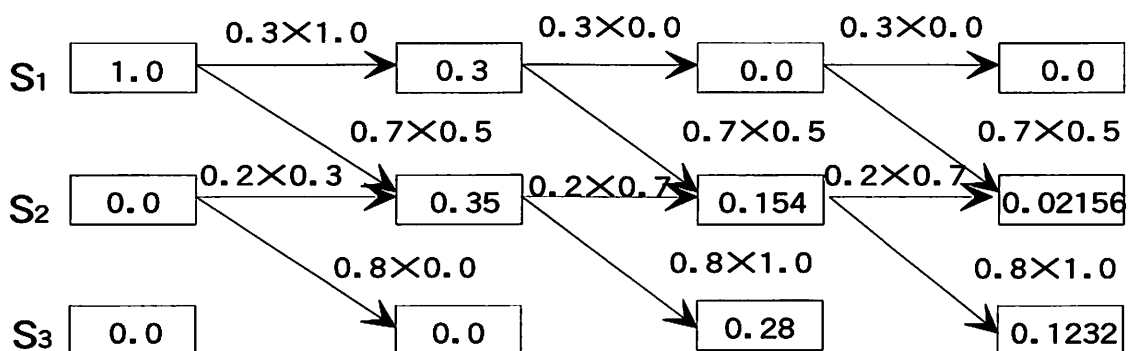


図 2.5: 図 2.3 の HMM における $P(abb)$ のトレリス上の計算

これより $P(abb) = \alpha(S_3, 3) = 0.1232$ となることが理解できるだろう。

2.4 Viterbi アルゴリズム

二つの単語 w_1 と w_2 に対応する HMM M_1 と M_2 を連結させて一つの HMM M_3 を構成したとしよう。この時、 M_1 の最終状態から M_2 の初期状態へナル遷移を設けて連結

するものとしよう（あるいは、 M_1 の最終状態を M_2 の初期状態に縮退させても良い）。 M_3 と入力文字列に対応する出力シンボル系列 y_1, y_2, \dots, y_T に対して前述のアルゴリズムで $P(y_1, y_2, \dots, y_T | M_3)$ を求めることができる。これは一般に、

$$P(y_1, y_2, \dots, y_T | M_3) = \sum_t P(y_1, y_2, \dots, y_t | M_1) \times P(y_{t+1}, y_{t+2}, \dots, y_T | M_2)$$

と表現できる。これは、 M_1 の最終状態からほかの状態への遷移先は M_2 の初期状態にかぎられる事から明らかである。上式の関係から、 $P(y_1, y_2, \dots, y_T | M_3)$ の算出の際には、 $w_1(M_1)$ にもっともよく対応するシンボル系列区間 y_1, y_2, \dots, y_t と $w_2(M_2)$ にもっともよく対応するシンボル系列区間 $y_{t+1}, y_{t+2}, \dots, y_T$ を決定する事はできない。ところが複合語の単語分割のためにこの最適な単語境界時点 t を知りたい。これは次のような定義

$$P'(y_1, y_2, \dots, y_T | M_3)$$

から求められる。

$$P'(y_1, y_2, \dots, y_T | M_3) = \max_t P(y_1, y_2, \dots, y_t | M_1) \times P(y_{t+1}, y_{t+2}, \dots, y_T | M_2)$$

上式の定義は近似的ではあるが妥当な定義とも考えられる。言い換えれば y_t を HMM M_3 の唯一の状態遷移に対応させたことになる。この考えをすべての y_t に拡張すれば、式

$$P(y_1, y_2, \dots, y_T) = \sum_{\mathbf{x}} \prod_i P(x_i | x_{i-1}) \cdot P(y_i | x_{i-1}, x_i)$$

の代わりに、

$$P''(y_1, y_2, \dots, y_T) = \max_{\mathbf{x}} \left\{ \prod_i P(x_i | x_{i-1}) \cdot P(Y_i | x_{i-1}, x_i) \right\}$$

を求めることになる。

これは式の $\{ \}$ 内の項を最大にする状態遷移系列上での確率で $P(y_1, y_2, \dots, y_T)$ を近似するものである。これによって、任意の y_t は唯一の状態遷移に対応付けられる。この状態遷移系列を最適パスと呼ぶ。

最適パスとこのパス上での確率を求めるためには、動的計画法を用いる。このアルゴリズムは最初 Viterbi によって提案された事から、Viterbi アルゴリズムと呼ばれている。 $Q(i, t)$ を $P''(y_1, y_2, \dots, y_t) = P(y_1, y_2, \dots, y_t, \{ \mathbf{x} \})$ を最大にするパス（状態遷移系列）、 $f(i, t)$ をその確率と定義すると Viterbi アルゴリズムは以下で与えられる。

Viterbi アルゴリズム

1. 初期化

すべての状態 i に対して $f(i, 0) = \pi_i$

2. $t = 1, 2, \dots, T$ に対して (3)、(4) を実行。

3. すべての状態 i に対して (4) を実行。

4. $\hat{i} = \operatorname{argmax}_j f(j, t) a_{ji}$

$$f(i, t) = \max\{f(\hat{j}, t-1) a_{\hat{j}i} \cdot b_{\hat{j}i}(y_t)\}$$

$$Q(i, t) = Q(\hat{j}, t-1) \otimes \hat{j}$$

5. $i = \operatorname{argmax}_{i \in F} f(i, T)$

$$P''(y_1, y_2, \dots, y_T) = P(y_1, y_2, \dots, y_T, Q(\hat{i}, T)) = f(i, T)$$

ここで、 \otimes は状態遷移系列と状態を連結し、新たな状態遷移系列をつくるオペレータである。最適状態遷移系列は（最適パス）は $Q(\hat{i}, T)$ で与えられる。HMM法による認識では、 $P(y)$ をもちいる代わりに、この $f(\hat{i}, T)$ を用いて最大値を与えるモデルのカテゴリーを認識結果とする方法もあり、同等の認識精度が得られている。

2.5 Viterbi アルゴリズムの例

形態素解析にもちいるHMMのモデル上では、ある状態からある状態へ遷移するさいに、あるシンボルを出力する。あるじょうたいへの遷移は存在する状態の数だけとることができ、それぞれ確率によって決まっている。初期状態から最終状態まで遷移し、観測できるのは出力シンボルのみである。HMMは出力シンボルが与えられても状態遷移系列は唯一にしか決まらない。さきの図 2.3 の例でHMMのモデルを考えてみる。このようなモデルがあったとき、最大の状態遷移確率と状態遷移系列を同時に求めるのがViterbi アルゴリズムである。つまり、このアルゴリズムによって状態遷移を最適にする状態遷移系列と、この系列上での確率を求めることができる。

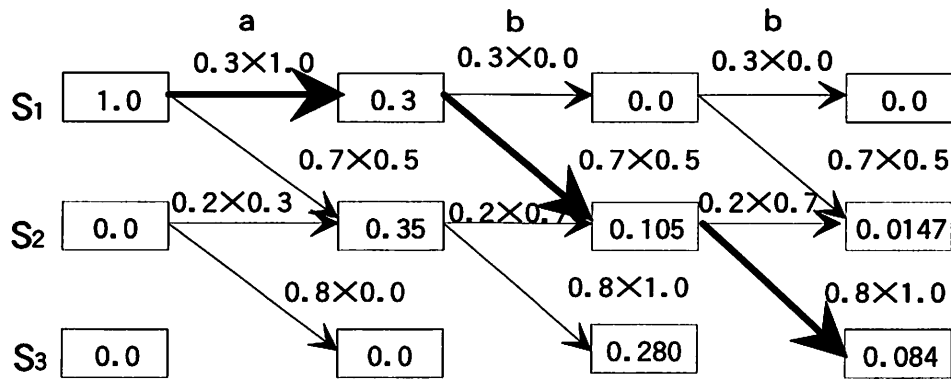


図 2.6: 図 2.3 の HMM における $P(abb)$ の Viterbi アルゴリズムによる最適パス算出例

第3章 本研究におけるHMMの設定

専門用語や一般語の各構成単語の出力過程を状態間の遷移と捉える。そして専門用語に各構成単語とその単語間の関係に専門用語らしさが内在していると考えれば、ある状態間を遷移して出力された複合語を専門用語と見なすことができる。本研究ではこの状態遷移を図3に示すオートマトンで表現した。 S_1 から出発し、 S_4 を通して最終状態 S_8 に遷移した単語列を専門用語と見なし、 S_1 から出発し、 S_7 を通して S_8 に遷移した単語列を一般語と見なす。出力シンボル列から遷移した状態列を推定するために図3.1のオートマトンを本研究のHMMとした。

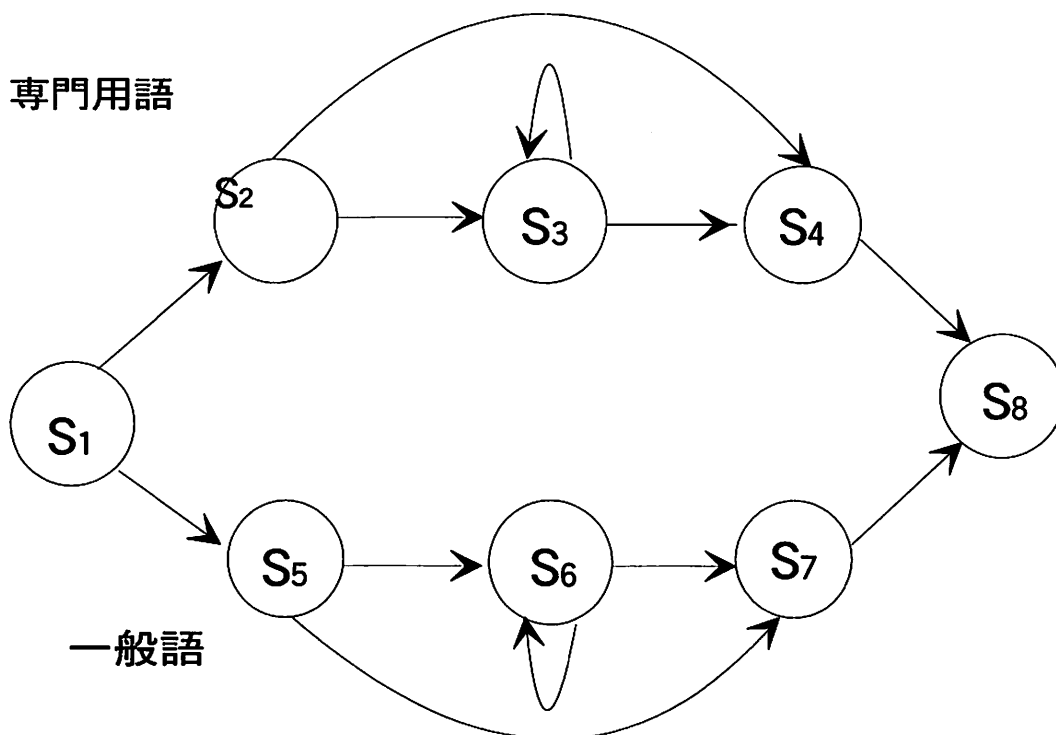


図 3.1: 専門用語判別のためのHMMモデル

3.1 各系列における確率設定

このモデルを活用するためには、HMMのパラメータ付与として訓練データを各状態に分類し、状態遷移確率と出力シンボルの条件付確率を割り当てなければならない。

以後、先ほどのモデルの上側の遷移系列を専門用語系列、下側を一般語系列と呼び、それぞれの確率割り当てを行なう。

3.1.1 状態遷移確率の割り当て

そのため、用意した一般語と専門用語を各5個、合計10個のグループに分類した。

専門用語

1. $S_2 \rightarrow S_4$

2語で構成された専門用語の先頭にある語句を置く。

例：

電子／メール → 電子

シリアル／ケーブル → シリアル

2. $S_2 \rightarrow S_3$

3語以上で構成された専門用語の先頭にある語句を置く。

例：

3語：空白／ロケータ／値 → 空白

4語：基本／形／リンク／制御 → 基本

3. $S_3 \rightarrow S_3$

4語以上で構成された専門用語の、先頭と最後とに挟まれた語句を置く。

例：

4語：待ち／行列／制御／テーブル → 行列

5語：制御／脚／書き／報告／集団 → 脚・書き

4. $S_3 \rightarrow S_4$

3語以上で構成された専門用語の、最後から1つ前にある語句を置く。

例：

3語：空白／ロケータ／値 \rightarrow ロケータ

4語：基本／形／リンク／制御 \rightarrow リンク

5語：制御／脚／書き／報告／集団 \rightarrow 報告

5. $S_4 \rightarrow S_8$

専門用語の最後にある語句を置く。

例：

電子／メール \rightarrow メール

ブロック／制御／単位 \rightarrow 単位

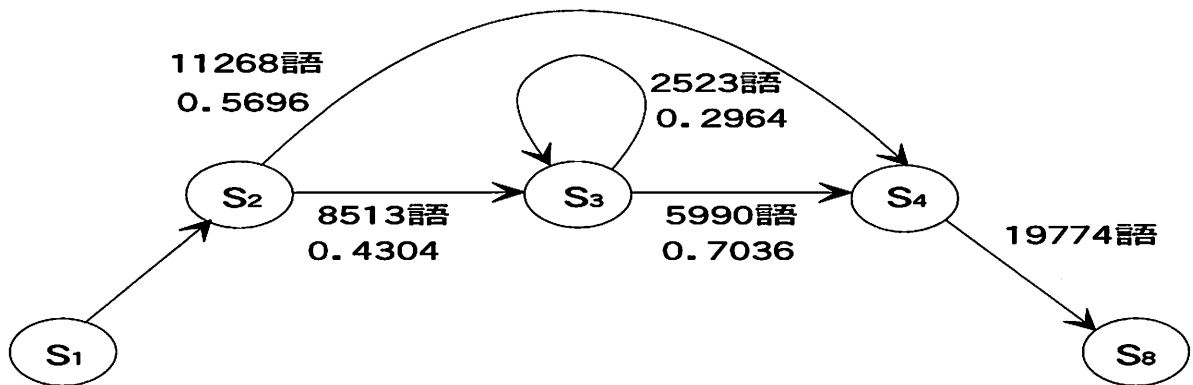


図 3.2: 専門用語系列の確率割り当て

一般語

6. $S_5 \rightarrow S_7$

2語で構成された一般語の先頭にある語句を置く。

7. $S_5 \rightarrow S_6$

3語以上で構成された一般語の先頭にある語句を置く。

8. $S_6 \rightarrow S_6$

4語以上で構成された一般語の、先頭と最後とに挟まれた語句を置く。

9. $S_6 \rightarrow S_7$

3語以上で構成された一般語の、最後から1つ前にある語句を置く。

10. $S_7 \rightarrow S_8$

一般語の最後にある語句を置く。

これら10のグループ分類から、各グループに属する語句の数を求めて、以下のように状態遷移確率を割り当てた。

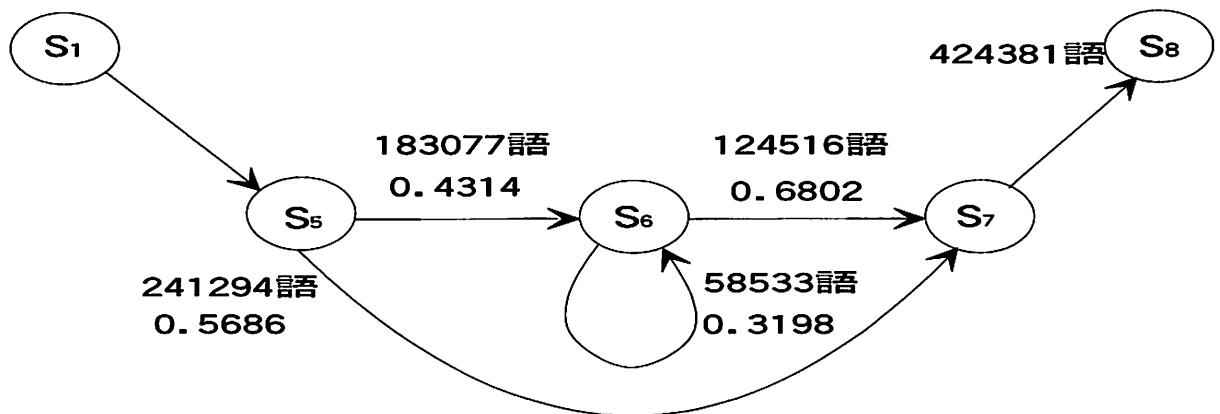


図 3.3: 一般語系列の確率割り当て

なお本研究では遷移 $S_1 \rightarrow S_2$ と $S_1 \rightarrow S_5$ さらに $S_4 \rightarrow S_8$ と $S_7 \rightarrow S_8$ の状態遷移確率を1とした。

3.1.2 出力シンボルの条件付き確率

次に各遷移状態での出力シンボルの確率を求める。これは全体に占める、その出力シンボルが出力される割合を示している。

以下に、実際に今回のモデルに使用した遷移 $S_2 \rightarrow S_4$ での出力シンボルの一部を示す。確率を求める前処理として、状態毎の含まれる語句の数をあらかじめ求めておいた

⋮

アウトオブキルター	1	0.000089
アウトライン	1	0.000089
アクション	3	0.000266
アクセス	19	0.001686
アクティブ	3	0.000266
アセンブラ	1	0.000089
アセンブリ	1	0.000089
アド	2	0.000177
アドレス	28	0.002485
アナログ	10	0.000887
アパチャ	1	0.000089
アパレル	1	0.000089
アン	2	0.000177
イエロー	1	0.000089
イオン	1	0.000089
イコール	1	0.000089
イタリック	1	0.000089
イナクティブ	1	0.000089
イニシャル	2	0.000177
イネーブル	1	0.000089
イベント	4	0.000355
イメージ	9	0.000799
イン	4	0.000355

インクリメンタル 1 0.000089

インクリメント 1 0.000089

インクワイアリ 2 0.000177

⋮

ここで語句の次の数字は、用意した専門用語集にいくつその語が登場したかという頻度である。全体の語数のうちのその語の占める割合が出力シンボル確率となる。

「インクワイアリ」を例にとると、これは専門用語集に2語含まれているので、全体の語数11268語より出力シンボル確率

$$2 \div 11268 = 0.000177$$

が得られる。

ただし、この専門用語集に該当のない語の出力シンボル確率は各遷移状態の該当数1のときの出力シンボル確率の半分と設定する。この設定を行なうことで、辞書に該当のない新用語を判別することが可能になる。

このような処理を上記で求めた10種の状態遷移ごとに行ない、出力シンボル確率を付与する。

以後このモデルと確率分布を基に判別処理を行なう。

3.2 複合語判別処理

抽出の対象である複合語をこのモデルに当てはめると以下の処理が行なわれる。「社内標準ソフト」を例として処理の流れを説明する。

これは／社内／標準／ソフト／の三語から構成されているので、考えられる状態遷移系列は $S_1S_2S_3S_4S_8$ と $S_1S_5S_6S_7S_8$ の2通りである。以下にそれぞれの系列毎の処理を詳細に説明する。

3.2.1 専門用語系列の遷移確率

最終的な確率は各遷移確率をかけあわせたもので表せる。つまり遷移系列 $S_1S_2S_3S_4S_8$ を通る確率は $S_1 \rightarrow S_2$, $S_2 \rightarrow S_3$, $S_3 \rightarrow S_4$, $S_4 \rightarrow S_8$ の4つの遷移確率をかけあわせたものである。

1. $S_1 \rightarrow S_2$

最初の遷移 $S_1 \rightarrow S_2$ の遷移確率は設定により 1 である。

$$P(S_1 \rightarrow S_2) = 1$$

2. $S_2 \rightarrow S_3$

専門用語の学習データを「社内」で検索した結果、該当語句は発見出来なかった。よって出力シンボル確率を一般語の該当数 1 の時の半分にあたる 0.000002 とおいた。

図 3.2 にあるようにこの時の遷移確率は 0.5696 なので、ここで求められる確率は

$$P(S_2 \rightarrow S_3) = 0.000001 \times 0.5696 = 0.000001$$

3. $S_3 \rightarrow S_4$

学習データより「標準」を検索。

標準 13 0.002170

結果 13 個の該当があり、出力シンボル確率 0.002170 を得た。また図 3.2 より遷移確率は 0.7036。これよりここでの確率は

$$P(S_3 \rightarrow S_4) = 0.002170 \times 0.7036 = 0.001527$$

4. $S_4 \rightarrow S_8$

学習データより「ソフト」を検索。

ソフト 4 0.001809

結果 4 個の該当があり、出力シンボル確率 0.001809 を得た。また $S_1 \rightarrow S_2$ と同様に遷移確率は 1 である。これよりここでの確率は

$$P(S_4 \rightarrow S_8) = 0.001809 \times 1 = 0.001809$$

この4つの確率から、専門用語系列 $S_1S_2S_3S_4S_8$ を通る確率 P_s は

$$P_s = 1 \times 0.000001 \times 0.001527 \times 0.001809 = 24 \times 10^{-13}$$

である。

3.2.2 一般語系列の遷移確率

専門用語の場合と同様に遷移系列 $S_1S_5S_6S_7S_8$ を通る確率は $S_1 \rightarrow S_5$, $S_5 \rightarrow S_6$, $S_6 \rightarrow S_7$, $S_7 \rightarrow S_8$ の4つの遷移確率をかけあわせたものである。

1. $S_1 \rightarrow S_5$

最初の遷移 $S_1 \rightarrow S_5$ の遷移確率は設定により1である。

$$P(S_1 \rightarrow S_5) = 1$$

2. $S_5 \rightarrow S_6$

学習データより「社内」を検索。

社内 35 0.000191

結果35個の該当があり、出力シンボル確率0.000191を得た。

図3.3にあるようにこの時の遷移確率は0.4314なので、ここで求められる確率は

$$P(S_5 \rightarrow S_6) = 0.000191 \times 0.4314 = 0.000082$$

3. $S_6 \rightarrow S_7$

学習データより「標準」を検索。

標準 21 0.000169

結果21個の該当があり、出力シンボル確率0.000169を得た。また図3.3より遷移確

率は 0.6802. これよりここでの確率は

$$P(S_6 \rightarrow S_7) = 0.000169 \times 0.6802 = 0.000115$$

4. $S_7 \rightarrow S_8$

学習データより「ソフト」を検索. 結果, 該当語句は発見出来なかった. よって出力シンボル確率を一般語の該当数 1 の時の半分にあたる 0.000019 とおいた.

また $S_1 \rightarrow S_5$ と同様に遷移確率は 1 である. これよりここでの確率は

$$P(S_7 \rightarrow S_8) = 0.000019 \times 1 = 0.000019$$

この 4 つの確率から, 専門用語系列 $S_1 S_5 S_6 S_7 S_8$ を通る確率 P_i は

$$P_i = 1 \times 0.000082 \times 0.000115 \times 0.000019 = 2 \times 10^{-13}$$

である.

こうして求めた各系列に対する遷移確率 P_s , P_i は

$$P_s = 24 \times 10^{-13}$$

$$P_i = 2 \times 10^{-13}$$

これにより, 両者を比較すると P_s の系列を通った確率がより高いことが分かる. こうした処理を経て「社内標準ソフト」は専門用語として分類される.

ここで注意すべきことは, 遷移 $S_1 \rightarrow S_2$ と $S_6 \rightarrow S_7$ における出力シンボルが該当数 0 の確率であることである. このことから, たとえ辞書にない新しい専門用語が現れても本手法で分類可能であることがいえる.

3.2.3 その他の場合

上記の例では3語によって構成された複合語の判別について説明した。その他の構成語数の語句の判別は、以下の遷移系列の確率比較によって行なう。

- 2語

$S_1S_2S_4S_8$ と $S_1S_5S_7S_8$

- 4語

$S_1S_2S_3S_3S_4S_8$ と $S_1S_5S_6S_6S_7S_8$

- 5語

$S_1S_2S_3S_3S_3S_4S_8$ と $S_1S_5S_6S_6S_6S_7S_8$

第4章 抽出実験

次に実際に文書に上記の処理を行ない，判別結果の評価を行なう．あらかじめ人手により検出しておいた専門用語のうちいくつを抽出できたのかを評価の基準とする．ここで用いる評価は次の値から行なう．

正解率 = 正解数 ÷ 抽出専門語数

再現率 = 正解数 ÷ 真の専門語数

f 値 = (2 × 正解率 × 再現率) ÷ (正解率 + 再現率)

以下に文書5個に対する実験結果を示す．この文書はすべて，情報分野に関する雑誌記事である．

表 4.1: 実験結果

対象	真専門	抽出	正解数	正解率	再現率	f 値
文書1	59	66	54	82%	92%	0.87
文書2	18	27	17	63%	94%	0.75
文書3	30	29	19	66%	63%	0.64
文書4	46	59	45	76%	98%	0.86
文書5	13	35	12	34%	92%	0.50

全体での f 値平均が 72% との結果が得られた．

またで挙げる 65 種類の新しい専門用語の抽出に成功した．

4.1 実例による処理説明

上記の実験結果を得るために行なった処理を、文書5を例にして説明する。

図1.1の流れ図を参考にしていきたい。

1. もとの文

まず抽出対象の文は以下のものである。なおここで表示したのは全体の一部である。

昨年来、常に刺激的な話題を提供し続けている、インターネット音楽配信の分野だが、先ごろ、またしても興味深い事件が起こった。

MP3による音楽配信の中核的サイトである『MP3.com』を、米レコード協会(RIAA)がニューヨーク連邦地裁に著作権侵害で提訴したのである。

問題にされたのは、『MP3.com』が2000年1月に同サイトのパーソナライズ・サービス『My.MP3.com』の一環としてスタートしたばかりの新サービス「Beam-it」と「Instant Listening Service」だ。

「Beam-it」は、ローカルのさまざまな音源を統括管理するいわゆるジュークボックス・ソフトをオンライン・アプリケーションとして拡張したようなサービス。手持ちのCDをパソコンのCD-ROMドライブに入れると、「Beam-it」のクライアント・アプリケーションは、「Beam-it」のサーバに、そのCDのデータが存在するかを見に行く。存在すれば、そのCDの楽曲リストが『My.MP3.com』の自分用のページに表示されるようになる。そしてユニークなのは、これ以降、CDをドライブから取り出しても、そのCDの音楽を『My.MP3.com』からのストリーミングで聴くことができるようになることだ。これによってユーザは、どこにいても、『My.MP3.com』にアクセスする環境があれば、自分の持っているCDを聴くことができることになる。

もうひとつの「Instant Listening Service」は、『MP3.com』と提携しているオンラインCDショップでCDを購入すると、そのCDが自宅に届く前に、CDの音楽をストリーミングで聴けるようになるというサービスだ。

音楽配信の事情を少しでも知っている人ならすぐわかると思うが、これはいずれも非常に面白いと同時に、なかなかキワどいサービスでもある。CDはリスナーが購入しているわけだから、著作権者の金銭面における権利を直接的に侵害はしていない。しかし、著作権者の許諾なく、音楽作品を配信していることには変わりない、と見れば見れないこともない。法解釈によってはクロの可能性もあるが、誰も実害を被らず、

ユーザにはベネフィットがあるのだからいいのではないか、という主張もあるだろう。逆に、レコード会社からすると、オフィシャルの音楽配信のためのデータベースを準備している段階で、勝手に自分とこの音源を配信されたのではたまらんといい思いがあるだろう。

2. 形態素解析

次にこの文を形態素解析によって分割分類を行なう。

昨年来	サクネンライ	昨年来	名詞-副詞可能	16 * 0 * 0
、	、	、	記号-読点	75 * 0 * 0
常に	ツネニ 常に	常に	副詞-一般	52 * 0 * 0
刺激	シゲキ 刺激	刺激	名詞-サ変接続	17 * 0 * 0
的	テキ 的	的	名詞-接尾-形容動詞語幹	31 * 0 * 0
な	ナ だ	だ	助動詞	70 特殊・ダ 48 体言接続 7
話題	ワダイ 話題	話題	名詞-一般	2 * 0 * 0
を	ヲ を	を	助詞-格助詞-一般	58 * 0 * 0
提供	テイキョー	提供	名詞-サ変接続	17 * 0 * 0
し	シ する	する	動詞-自立	44 サ変・スル 3 連用形 8
続け	ツズケ 続ける	続ける	動詞-非自立	45 一段 6 連用形 5
て	テ て	て	助詞-接続助詞	61 * 0 * 0
いる	イル いる	いる	動詞-非自立	45 一段 6 基本形 2
、	、	、	記号-読点	75 * 0 * 0
インターネット	インターネット	インターネット	名詞-一般	2 * 0 * 0
音楽	オンガク	音楽	名詞-一般	2 * 0 * 0
配信	ハイシン	配信	名詞-サ変接続	17 * 0 * 0
の	ノ の	の	助詞-連体化	67 * 0 * 0
分野	ブンヤ 分野	分野	名詞-一般	2 * 0 * 0
EOS				
だが	ダガ だが	だが	接続詞	55 * 0 * 0
、	、	、	記号-読点	75 * 0 * 0
先ごろ	サキゴロ	先ごろ	名詞-副詞可能	16 * 0 * 0
、	、	、	記号-読点	75 * 0 * 0
また	マタ また	また	副詞-一般	52 * 0 * 0
し	シ する	する	動詞-自立	44 サ変・スル 3 連用形 8
て	テ て	て	助詞-接続助詞	61 * 0 * 0
も	モ も	も	助詞-係助詞	62 * 0 * 0

興味深い	キョーミブカイ	興味深い	形容詞-自立 48	形容詞-アウオ段 42	基
本形 2					
事件	ジケン	事件	名詞-一般	2 * 0 * 0	
が	ガ	が	助詞-格助詞-一般	58 * 0 * 0	
起こっ	オコッ	起こる	動詞-自立 44	五段・ラ行 17	連用タ接続 8
た	タ	た	助動詞 70	特殊・タ 47	基本形 2
。	。	。	記号-句点	74 * 0 * 0	
EOS					
			記号-空白	76 * 0 * 0	
MP3	未知語	MP3	名詞-サ変接続	17 * 0 * 0	
による	ニヨル	による	助詞-格助詞-連語	60 * 0 * 0	
音楽	オンガク	音楽	名詞-一般	2 * 0 * 0	
配信	ハイシン	配信	名詞-サ変接続	17 * 0 * 0	
の	ノ	の	助詞-連体化	67 * 0 * 0	
中核	チューカク	中核	名詞-一般	2 * 0 * 0	
的	テキ	的	名詞-接尾-形容動詞語幹	31 * 0 * 0	
サイト	サイト	サイト	名詞-一般	2 * 0 * 0	
で	デ	だ	助動詞 70	特殊・ダ 48	連用形 5
ある	アル	ある	助動詞 70	五段・ラ行アル 18	基本形 2
「	「	「	記号-括弧開	77 * 0 * 0	
MP3.com	未知語	MP3.com	名詞-サ変接続	17 * 0 * 0	
」	」	」	記号-括弧閉	78 * 0 * 0	
を	ヲ	を	助詞-格助詞-一般	58 * 0 * 0	
、	、	、	記号-読点	75 * 0 * 0	
米	{アメリカ/ペイ}	米	名詞-固有名詞-地域-国	12 * 0 * 0	
レコード	レコード	レコード	名詞-一般	2 * 0 * 0	
協会	キョーカイ	協会	名詞-一般	2 * 0 * 0	
(((記号-括弧開	77 * 0 * 0	
RIAA	未知語	RIAA	名詞-サ変接続	17 * 0 * 0	
)))	記号-括弧閉	78 * 0 * 0	
EOS					
が	ガ	が	接続詞	55 * 0 * 0	
ニューヨーク	ニューヨーク	ニューヨーク	名詞-固有名詞-地域-一般	11 * 0 * 0	
連邦	レンポー	連邦	名詞-一般	2 * 0 * 0	
地裁	チサイ	地裁	名詞-一般	2 * 0 * 0	
に	ニ	に	助詞-格助詞-一般	58 * 0 * 0	
著作	チョサク	著作	名詞-サ変接続	17 * 0 * 0	

権	ケン	権	名詞-接尾-一般 28 * 0 * 0
侵害	シンガイ		侵害 名詞-サ変接続 17 * 0 * 0
で	デ	で	助詞-格助詞-一般 58 * 0 * 0
提訴	テイソ	提訴	名詞-サ変接続 17 * 0 * 0
し	シ	する	動詞-自立 44 サ変・スル 3 連用形 8
た	タ	た	助動詞 70 特殊・タ 47 基本形 2
の	ノ	の	名詞-非自立-一般 21 * 0 * 0
で	デ	だ	助動詞 70 特殊・ダ 48 連用形 5
ある	アル	ある	助動詞 70 五段・ラ行アル 18 基本形 2
。	。	。	記号-句点 74 * 0 * 0

EOS

⋮

3. 複合語抽出処理

次に形態素解析の結果から、複合語の部分のみを抽出する。

刺激/的

インターネット/音楽/配信

音楽/配信

中核/的/サイト

米/レコード/協会

ニューヨーク/連邦/地裁

著作/権/侵害

同/サイト

新/サービス

統括/管理

CD-ROM/ドライブ

楽曲/リスト

自分/用

CD/ショップ

著作/権/者

金銭/面

音楽/作品

法/解釈

可能/性

レコード/会社

4. 専門用語判別プログラム

最後に専門用語判別プログラムによって遷移確率を求め判断を下す.

$$0.000001 * 0.039349 * 100000 = 0.002241$$

$$0.000018 * 0.000019 * 100000 = 0.000033$$

刺激/的 0.002241 0.000033 専門

$$0.000001 * 0.000118 * 0.000452 * 10000000 = 0.000000$$

$$0.000009 * 0.000003 * 0.000404 * 10000000 = 0.000000$$

インターネット/音楽/配信 0.000000 0.000000 専門

$$0.000101 * 0.010925 * 0.000905 * 10000000 = 0.009998$$

$$0.000045 * 0.000003 * 0.000183 * 10000000 = 0.000000$$

中核/的/サイト 0.009998 0.000000 専門

$$0.000001 * 0.001644 * 0.002261 * 10000000 = 0.000032$$

$$0.000001 * 0.000060 * 0.000019 * 10000000 = 0.000000$$

米/レコード/協会 0.000032 0.000000 専門

$$0.000001 * 0.000013 * 0.000019 * 10000000 = 0.000000$$

$$0.000193 * 0.000328 * 0.003303 * 10000000 = 0.002093$$

ニューヨーク/連邦/地裁 0.000000 0.002093 一般

$$0.000101 * 0.000470 * 0.000019 * 10000000 = 0.000009$$

$$0.000080 * 0.000003 * 0.000771 * 10000000 = 0.000002$$

著作/権/侵害 0.000009 0.000002 専門

$$0.000303 * 0.000905 * 100000 = 0.027426$$

$$0.000001 * 0.000183 * 100000 = 0.000010$$

同/サイト 0.027426 0.000010 専門

$$0.000051 * 0.028494 * 100000 = 0.144458$$

$$0.000001 * 0.000019 * 100000 = 0.000001$$

新/サービス 0.144458 0.000001 専門

0.000001 * 0.036183 * 100000 = 0.002061

0.000018 * 0.000019 * 100000 = 0.000033

統括/管理 0.002061 0.000033 専門

0.000001 * 0.002261 * 100000 = 0.000129

0.000001 * 0.000734 * 100000 = 0.000042

CD-RDM/ドライブ 0.000129 0.000042 専門

0.000001 * 0.021710 * 100000 = 0.001237

0.000008 * 0.000019 * 100000 = 0.000015

楽曲/リスト 0.001237 0.000015 専門

0.000001 * 0.003618 * 100000 = 0.000206

0.000038 * 0.000019 * 100000 = 0.000071

自分/用 0.000206 0.000071 専門

0.000001 * 0.000452 * 100000 = 0.000026

0.000001 * 0.002606 * 100000 = 0.000148

CD/ショップ 0.000026 0.000148 一般

0.000101 * 0.000452 * 100000 = 0.004557

0.000001 * 0.000404 * 100000 = 0.000023

音楽/配信 0.004557 0.000023 専門

0.000101 * 0.000470 * 0.017187 * 10000000 = 0.008170

0.000080 * 0.000003 * 0.000019 * 10000000 = 0.000000

著作/権/者 0.008170 0.000000 専門

0.000001 * 0.004071 * 100000 = 0.000232

0.000044 * 0.000019 * 100000 = 0.000084

金銭/面 0.000232 0.000084 専門

0.000101 * 0.000019 * 100000 = 0.000192

0.000001 * 0.000019 * 100000 = 0.000001

音楽/作品 0.000192 0.000001 専門

0.000253 * 0.000452 * 100000 = 0.011432
0.000074 * 0.001798 * 100000 = 0.013290
法/解釈 0.011432 0.013290 一般

0.000152 * 0.000019 * 100000 = 0.000288
0.000001 * 0.000019 * 100000 = 0.000002
可能/性 0.000288 0.000002 専門

0.000152 * 0.003166 * 100000 = 0.047972
0.000038 * 0.000019 * 100000 = 0.000071
レコード/会社 0.047972 0.000071 専門

以上のような処理を経て、5つの文書の専門用語抽出を行なった。

4.2 詳細

それぞれの文書ごとのさらに詳しい実験結果を記す。ここで言う「誤抽出」は誤って専門用語として抽出した一般語、「未抽出」は専門用語として抽出できなかった語である。「新専門語」は今回用いたデータに該当例のない語を含む専門用語である。

なお出力の完全なリストは巻末に付与したので、そちらを参考していただきたい。

1. 文書 1

- 内容

パソコンの性能向上による、企業のリプレースへの対応。

- 誤抽出

ご購入/ご利用/ご了承/グループ販売会社/営業所/営業部隊
各企業/関係者/参加企業/資産管理/実用的/障害監視
新パソコン/生産性/製品計画/製品品質/会社展開/操作方法
総コスト/測定調査/大規模/大量導入/値自体/電力会社
日経コンピュータ/日本マクドナルド/販売戦略/部品代

本格的／本格導入／本記事／利用者

- 未抽出

グループウェア／システム所有／データ交換／ユーザー企業
運用コスト／業務アプリケーション／次期版／社内システム
添付ファイル／リソース不足／文書ファイル

- 新専門語

モバイル用ノート／マーケティング部門／98形式
リプレース作業／運用コスト削減／高スペック／社内標準ソフト
処理性能不足／情報化／新システム／性能不足／生産性／大量リプレース

2. 文書2

- 内容

インターネットを利用した、オンライン証券取り引きサービス。

- 誤抽出

外資系／共同出資者／自由化／主要株主／住友グループ
証券会社／世界初

- 未抽出

オンライン証券

- 新専門語

インターネット金融戦略／インターネット取り引き／オンライン取引
オンライン専業／ネットワーク時代／株式関連フォーラム

3. 文書3

- 内容

異なるシステム間を連携するためのE A I ツールの説明.

- 誤抽出

企業内／経営環境／取引先／新製品／全世界／部分的

- 未抽出

システム間／システム統合／システム同士／システム連携

データ項目／ミドルウェア／ミドルウェア製品

ユーザー企業／既存システム／連携プログラム

- 新専門語

E R P パッケージ／インターフェイス機能／サプライチェーン管理

システム間連携／データ変換機能／マッピング機能

メインフレーム／稼働実績／会計システム／企業間システム連携

基幹系システム／生産管理システム／通信機能／適用例

販売管理システム／連携用プログラム

4. 文書 4

- 内容

電子ブックの今後の可能性.

- 誤抽出

2 時間未満／8 時間／可能性／危険性／具体的

研究機関／商業的／消極的／制限時間／先駆者

対応作／対処法／大企業／販売価格

- 未抽出

D H C P サーバー／キーノートスピーチ／ユーザー自身

- 新専門語

DHCPサーバー間/DHCPサーバー側/IPアドレス

ISDN回線/LAN経由/Mac側/PPP接続

USBポート/スタンド型/ダイヤルアップルータ自体

ハンディ端末/ユーザー側/一般ユーザー/貸し出し作業

通常IPアド/電子ブックリーダー/同アップデート

日本語OS/連続表示時間

5. 文書5

- 内容

インターネットを利用した音楽配信ビジネス。

- 誤抽出

1500円分/ゲリラ的/レコード会社/音楽ソフト

音楽ファン/音楽作品/可能性/確信犯的/楽曲リスト

金銭面/高品質/差別化要素/最新情報/刺激的

自分用/新サービス/世界的/選択眼/著作権者

著作権侵害/米レコード協会/本格的/有線放送

- 未抽出

PCユーカラ

- 新専門語

CD-ROMドライブ/インターネット音楽配信

音楽コンテンツ/音楽ポータル/中核的网站

第5章 考察

上記実験では再現率に優れた値が得られたが反対に正解率は低かった。以下でその原因を考える。

5.1 低正解率の原因

今回の実験では平均64.2%の正解率が得られた。ここで正解率の低い原因として考えられるのは専門用語と一般語の数の差である。用意した専門用語が一般語より圧倒的に少ないため、一語あたりの出力シンボル確率の重みに大きな差がでてしまう。

そのため、誤抽出例のような「~的」「~面」といった一般語にも多くある語も、前後の語句の出力シンボル確率が低いため専門用語として分類してしまう。

例をいくつか用いて説明する。

商業／的

$$\text{専門 } 0.000001 \times 0.039349 \times 100000 = 0.002241$$

$$\text{一般 } 0.000094 \times 0.000019 \times 100000 = 0.000178$$

商業／的 = 専門用語

金銭／面

$$\text{専門 } 0.000001 \times 0.004071 \times 100000 = 0.000232$$

$$\text{一般 } 0.000044 \times 0.000019 \times 100000 = 0.000084$$

金銭／面 = 専門用語

手／入力

$$\text{専門 } 0.000051 \times 0.022162 \times 100000 = 0.112356$$

$$\text{一般 } 0.000063 \times 0.000697 \times 100000 = 0.004399$$

手／入力＝専門用語

これらの学習データでの語句の該当数を見ると

商業 = 専門0個，一般70個

的 = 専門87個，一般0個

金銭 = 専門0個，一般33個

面 = 専門9個，一般0個

手 = 専門1個，一般47個

入力 = 専門49個，一般19個

「一般」「金銭」「手」といった語句は一般語の学習データに多く該当が見られるがその該当数が結果に反映されずらいということがいえる。

その理由はデータの数の差により専門用語の重みが増しているため，1該当ごとに割り当てられる出力シンボル確率に大きな差が出てしまうことである。

実際に遷移 $S_4 \rightarrow S_8$ (専門) と遷移 $S_7 \rightarrow S_8$ (一般) を比較すると，1該当あたりに割り当てられる出力シンボル確率は(専門) 0.000452 に対し(一般) 0.000037 と約12倍もの差がある。

この対策としては，専門用語の学習データをより多くし，一般語との数差をなくすことが考えられる。それにより出力シンボル確率の最適化が計れるだろう。

5.2 再現率について

一方で再現率は平均87.8%と高かったが，なかには「～ファイル」「システム～」のように，用意した専門用語に該当がなかったため，出力シンボル確率が低下し抽出できなかった語句もあった。

いくつか例を示す。

システム／統合

専門 $0.000001 \times 0.001809 \times 100000 = 0.000103$

一般 $0.000057 \times 0.001798 \times 100000 = 0.010326$

システム／統合＝一般語

データ／項目

専門 $0.000001 \times 0.030303 \times 100000 = 0.001726$

一般 $0.000089 \times 0.003156 \times 100000 = 0.027994$

データ／項目＝一般語

文書／ファイル

専門 $0.000960 \times 0.000019 \times 100000 = 0.001825$

一般 $0.000057 \times 0.000550 \times 100000 = 0.002877$

文書／ファイル＝一般語

これらの学習データでの語句の該当数を見ると

システム = 専門0個，一般43個

統合 = 専門4個，一般49個

データ = 専門0個，一般66個

項目 = 専門67個，一般86個

文書 = 専門19個，一般39個

ファイル = 専門0個，一般15個

未抽出の原因はこのように，学習データに該当がなかったため出力シンボル確率が低下し，結果的な遷移確率が下がったことによる。

これより，低正解率の原因と同様に専門用語をより多く用意し，出力シンボル確率を最適化することにより，結果の改善が見込まれる。

第6章 おわりに

本研究ではHMMを用いた専門用語抽出を行った。

実験結果からHMMによる情報分野での専門用語の抽出は有効であるといえる。特に辞書にない新しい専門用語を、抽出可能であることは大きな利点である。

さらに他分野の専門用語も、同様の手法を用いることにより自動抽出が可能になると予想できる。

今後より多くの専門用語を用意することにより、学習データが充実しさらなる信頼性が得られるだろう。

謝辞

本研究の遂行および論文の作成において多大なご助言および御指導を賜った 新納 浩幸 教官（茨城大学工学部システム工学科）に深い感謝の意を表します。

また本研究で利用したコーパスおよび評価文は、コンピュータ用語辞典 (CD-ROM) と毎日新聞 CD-ROM '94 版から得ています。利用を許可していただいた日外アソシエーツ社と毎日新聞社に深く感謝します。

最後にシステム工学科計算機応用講座の教官の方々にも深く感謝致します。

関連図書

- [1] 日本語学-特集「専門用語」-, 明治書院, Vol. 16, No. 2, pp.4-69 (1997).
- [2] 吉岡真治, 岡田真穂, 影浦峯, 小山照夫: 専門用語抽出・解析処理を考慮したコーパスの作成, 情報処理学会, 情報基礎研究会, 53-6, pp.41-47 (1999).
- [3] コンピュータ用語辞典, 日外アソシエーツ (1997).
- [4] 電子版毎日新聞94年度版, 毎日新聞社 (1994).
- [5] 日経コンピュータ9月13日号, Nikkei Business Publications(1999).
- [6] J. R. キンラン: AIによるデータ解析, トッパン (1993).
- [7] 中川聖一郎: “確率モデルによる音声認識”, 電子情報通信学会 (1988).

実験出力データ

文書1

本文

パソコンの大量リプレース

21世紀に見る、今踏み切る理由と乗り切る工夫

1995年前後、パソコン一人1台ブームやWindows 95の登場によって、パソコンを大量に導入する企業が相次いだ。そして今、“老朽化”したこれらのパソコンをリプレースする必要に迫られている企業は少なくないはずだ。だが、パソコンをリプレースするには、手間もコストもかかる。情報システム部門にとっては、できればやりたくない作業だ。しかし、リプレースの“大波”は否応なしにやってくる。各企業はさまざまな工夫を凝らし、パソコンの大量リプレースを乗り切ろうとしている。最近、リプレースを実施した大手企業21社の事例に基づき、「なぜ今リプレースに踏み切ったのか」、「リプレースで発生する様々な作業をどう上手く乗り切ったのか」の疑問に答える。

玉置 亮太, 中條 将典

本記事は日経コンピュータ99年9月13日号からの抜粋です。そのため図や表が一部割愛されていることをあらかじめご了承ください。なお本号のご購入はバックナンバー、または日経コンピュータの定期ご購読をご利用ください。

ユーザー企業を「パソコンの大量リプレース」という“大波”が襲いつつある（図1）。1994年から1995年にかけて企業に大量導入されたパソコンが、1999年から2000年にかけて一斉に「リプレース時期」を迎えるからだ。

古いパソコンを撤去して新しいパソコンを導入することは、企業にとって必ずしもよいことばかりではない。むしろ、「パソコンのリプレース作業には手間もコストもかかる。一度導入した機種はできるだけ長く使い続けたい」（日本マクドナルドの江口博章経営情報本部情報システム部統括マネージャー）というのが、情報システム部門の偽らざる本音だろう。

しかしリプレースの大波は、否応なしにやってくる。ハードの性能向上に歩調を合わせるかのように、アプリケーションを稼働させるのに必要な動作環境の仕様はどんどん高くなっていく。実際、最近のアプリケーションの中には、4～5年前のパソコンでは満足に動作させることが難しいものも珍しくない。

現在使っているパソコンに主記憶やハードディスクを増設してしのぐ手もあるだろう。だが、「増設の手間やコスト、部品代を考えると、結局新しいパソコンを買うのと同じくらい費用がかかる」（三井建設の矢野口薫正 経営企画部情報化推進室長）。

最新のアプリケーションが使えないということだけなら、利用者にガマンさせればすむかもしれない。だが“旧式”のパソコンを使い続けることにより、利用者の業務の生産性が低下するとなるとそうもいかない。例えば、Office 95以前のオフィス・ソフトを使っている企業の社員が、電子メールなどを使った社外とのやり取りでOffice 97形式の文書ファイルを受け取ると、そのたびに自分が読める形式に変換する手間を強いられる。ハードが老朽化して故障発生率が増加すれば、当然業務にも支障が出る。

最近、こういった状況を受けて、パソコンの大量リプレースに踏み切る企業が相次いでいる。本誌は、1995年前後にパソコンを大量に導入して、ここ1年ほどの間にそれらのパソコンをリプレースした企業21社取材した（122～125ページの表1）。21社の事例を見ると、各社がリプレースを決断した理由として挙げたことは大きく三つある。（1）新システムを稼働させた、（2）社外とのデータ交換に支障があった、（3）運用コストが増加した、である。もちろん、この三つのうちの複数をリプレースの理由として挙げる企業も少なくない。

理由1：新システムを稼働

ユーザー企業がパソコンのリプレースに踏み切るきっかけの一つに、新システムの稼働がある。新たに稼働させるクライアント/サーバー・システムが現在導入しているパソコンからは使えなかったり、動作しても低速で実用に耐えない場合にはリプレースは不可欠だ。

R/3 導入を機にリプレースを実施

自動車用タイヤ大手の東洋ゴム工業は、1998年10月から1999年4月にかけて、同社とグループ販売会社のWindows 3.1パソコン計800台をすべてWindows 95パソコンにリプレースした(図2)。

同社は1997年から、経営のスピード・アップや製品品質の向上を目指して、矢継ぎ早に社内システムを刷新している。新システムの核は、ERPパッケージ(統合業務パッケージ)の「R/3」である。1999年4月に人事と会計モジュールの運用を開始し、その後も生産管理モジュールなどを相次いで稼働させる予定である。

これに先だって1997年4月には、グループウェア・ソフトのExchange Serverの利用を開始。さらに製品計画や販売戦略などの立案に役立てるためにデータ・ウェアハウスを構築し、分析/検索ソフトであるImpromptuとPowerPlayも運用している(詳細は本誌1998年11月23日号140ページ「トレンド」を参照)。

同社がシステム刷新計画の一環としてパソコンのリプレースを検討し始めたのは1998年1月。当時の社内では、1995年ごろに導入したWindows 3.1パソコンがまだ大量に使われていた。そのスペックは、プロセッサが動作周波数25MHzまたは33MHzのi386、主記憶8MB、ハードディスク容量340MBというものである。「新システムが稼働すれば、R/3のクライアントとして使うWWWブラウザ、Exchangeのクライアント、データ・ウェアハウス用の分析ソフトなど複数のアプリケーションを1台のパソコンで利用ようになる。これらの新システムを快適に使うには、まったく性能不足だった」(金居利夫 経営企画部システム担当部長)。

しかも旧式のパソコンは、システム部門やマーケティング部門など、利用部門の中でパソコンをより“ヘビー”に使う部署に集中していた。東洋ゴムがパソコンを本格導入し始めた1995年当時、こうした部署に真っ先に導入したからである。そのため、「パソコンをよく使う部署ほど旧式のパソコンを使っているという“ねじれ”が発生していた」(市川貴史 経営企画部システム担当部長代理)。

新パソコンは「新システムがすべて実用的な速度で動作することを確認したうえで、最低でも今後2～3年は使い続けられるよう、1998年当時としては比較的高スペックの機種を選んだ」(市川部長代理)。

リソース不足でフリーズが多発

アサヒビールは、パソコンのリプレースを決断した理由として、処理性能不足に加えて、業務アプリケーションの開発に支障が出ていたことを挙げる(図3)。

同社がWindows 3.1パソコンを導入し始めたのは1994年のこと。5年または4年のリース契約を結んでいた。しかし導入から3年ほどが経過した1997年10月には、「リース切れまで待てない」（尾関博 総合品質本部業務高度化推進部副主任）と判断して、大規模なリプレースに踏み切った。1999年3月までのおよそ1年半をかけて、計2000台のWindows 3.1パソコンをすべてWindows NT 3.51パソコンにリプレースした。

同社がリプレースを検討し始めたのは、業務アプリケーションが動作するWindows 3.1パソコンが「毎日のようにフリーズしていた」（尾関副主任）からである。当時使っていたパソコンのスペックは、プロセッサが動作周波数75MHzのi486、主記憶16MB、ハードディスクは最大でも800MBというもの。同社が開発した業務アプリケーションだけでなく、Word 6.0やExcel 5.0といったオフィス・アプリケーションを使っている時にも、リソース不足に起因するとみられるフリーズは多発した。

理由2：データ交換に支障

社外とのデータ交換に関する苦勞をリプレースの理由に挙げる企業もある。最も多いのは、旧式のパソコンで使っているアプリケーションでは、社外からの電子メールに添付されたファイルを開けないというものだ。

こうした問題を解決するには、アプリケーションをバージョンアップするという選択肢もあり得る。しかし旧式のパソコンは、Office 97やOffice 2000といった最近主流のアプリケーションを動作させるには“力不足”であることが多い。

メールで受け取ったファイルを読めない

中国電力は1994年からノート・パソコンの大量導入を開始、1997年には約1万人の社員の「パソコン一人1台体制」を確立した。1997年当時の、パソコンの内訳はWindows 3.1パソコンが約5700台、Windows 95パソコンが約4700台である。

中国電力は1999年末までに社内のWindows 3.1パソコンを一掃する。まず1998年7月から同年11月にかけて、2000台のWindows 3.1パソコンを、Windows 95搭載のノート・パソコンにリプレース。引き続き1999年8月から年末にかけて3700台のノート・パソコンを、同様にリプレースする（図4）。

同社をリプレースに踏み切らせた理由の一つは、社外とのデータ交換に関する利用部門からの不満の声であった。中国電力は「全社の方針として、社内で使うアプリケーションは統一している」（情報通信部情報技術担当の浜田浩和氏）。したがって、オフィス・ソフトは、Windows 3.1パソコンでも稼働するOffice 4.2に統一している。

Office 4.2はもはや旧式のソフトではあるが、社内同士で文書をやり取りするにはまったく問題はない。だが、社外との間となると話は別だ。他の電力会社の関係者などと電子メールでやり取りする際には、添付ファイルはOffice 4.2形式のものにしてもらうようにいちいち依頼する必要があった。

利用者とヘルプデスクに負担がかかる

1999年3月に社内1500台のWindows 3.1パソコンをWindows 98にリプレースする作業に着手したライオンも、大きな理由の一つとして社外とのデータ交換を挙げる。同社では「特に量販店向けや卸売業者向けの営業部隊など社外とやり取りする部署を中心に、不満を訴える声が多かった」(吉本先知 統合システム部長)。

ライオンは1997年末にはすでに社員一人に1台のパソコンを配備し、「社内標準ソフト」としてワープロはDTP(デスクトップ・パブリッシング)ソフトのPRESBOX、表計算ソフトはExcel 5.0を指定していた。社外とのデータ交換を頻繁に行う部署については、Word 5.0をインストールすることも許可していたが、1998年にはすでに、世間ではWord 5.0は“少数派”になっていた。

社外とのやり取りでWord 95や同98の添付ファイルを受け取った利用者は、社内のヘルプデスク担当者に送ってWord 5.0で読める形式に変換して送り返してもらっていた(図5)。だが、電子メールでのやり取りが頻繁になるにつれ、ヘルプデスクにファイル変換を依頼する手間が無視できなくなってきた。

ヘルプデスク担当者の負担も大きかった。ユーザーから受け取ったWord 95/98形式のファイルをいちいち、Word 98で読み込んでWord 5.0形式にセーブし直して送り返していたからである。

加えて、社内で利用するほとんどのアプリケーションは、次期版でWindows 3.1版を出荷しないことが分かった。そこでライオンは、リース切れになる機種から順に、Office 97を搭載したWindows 98パソコンにリプレースすることを決めた。

理由3：運用コストが増加

ユーザー企業がパソコンのリプレースに踏み切る理由の三つ目として挙げるのは、パソコン運用のコストが増加してきたことである。

運用のコストとしては、情報システム部門による技術サポートや管理ツール導入といった目に見えるもの以外に、利用部門が行う“隠れた運用のコスト”が大きい。これは、利用部門がトラブルの

原因究明や復旧に費やす時間や、パソコンの操作方法について同僚を手助けすることなどに費やす時間をコストに換算したものである。パソコンにトラブルが発生したことによって利用者の業務に支障が出れば、それも含まれる。

TCO測定調査がきっかけ

味の素ゼネラルフーズ（AGF、東京都品川区）は、1997年に導入した本社のWindows NT4.0パソコン150台と、1996年に導入した営業所のWindows 95パソコン360台の合計510台を、Windows NT4.0パソコンにリプレースする。作業は1999年10月から11月にかけて実施し、リプレースされたWindows NT4.0パソコンのうち一部は、他の部署に流用する。AGFは社内のTCO（システム所有の総コスト）の測定調査と、利用部門へのアンケートを実施した結果、運用コストの大きさに気づき、リプレースに踏み切ることを決めた（図6）。

「TCOの値自体は、測定調査に参加した参加企業の平均を下回っていた」（井上博志情報システムセンター主査）。だが「利用部門による運用のコストが、AGFのTCO全体の50%を占めており、思っていた以上に大きいことが分かった」（同）。

さまざまな機種が導入され管理が困難に

旭化成工業も、運用コスト削減を目標にリプレースを決断した。同社は1999年4月から2000年3月にかけて、1995年に導入したWindows 3.1パソコン3000台をすべてWindows 95パソコンにリプレースする。

旭化成がパソコンを本格的に導入し始めたのは1995年にさかのぼる。当時は社内LANの整備や電子メールの全社展開など、社内の情報化が急速に進みつつあった。導入するパソコンについては特に制限はなく、そのつど機種を選んでいったという。

その結果、“まだら模様”のようにさまざまな機種のパソコンが導入されていった。「機種がバラバラの環境では、運用コストがかさむことは明らかだった。障害が起こった際に、素早く原因を特定して復旧処置をすることが難しいからだ」（上野史彦 経営計画管理部業務革新推進室長）。そこで旭化成は1999年4月、1995年に導入したパソコンのリース契約が切れるのを機に、大量リプレースに踏み切った。

旭化成は「運用コスト削減」というリプレースの狙いを徹底するため、リプレースに合わせて二つの対策を実施した。一つは今後の無秩序なパソコンの導入を抑制するために、デスクトップ、ノート、モバイル用ノートのそれぞれについて「標準パソコン」を1機種ずつ定め、同時に購買窓口をイントラネットの「社内標準パソコン購入用ページ」に一本化すること。もう一つが、資産管理や障害監視の体制を整備するために、運用管理ツールを導入することである。

続きは日経コンピュータ99年9月13日号をお読み下さい。この号のご購入はバックナンバー、
または日経コンピュータの定期ご購読をご利用ください。

このホームページに掲載されている記事・写真・図表などの無断転載を禁じます。

判別結果

0.000001 * 0.020805 * 100000 = 0.001185

0.000003 * 0.000019 * 100000 = 0.000005

9 8 /形式 0.001185 0.000005 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.000019 * 100000 = 0.000001

ご/購入 0.000001 0.000001 専門

0.000001 * 0.003166 * 100000 = 0.000180

0.000001 * 0.000019 * 100000 = 0.000001

ご/利用 0.000180 0.000001 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.000477 * 100000 = 0.000027

ご/了承 0.000001 0.000027 一般

0.000001 * 0.000905 * 100000 = 0.000052

0.000007 * 0.002459 * 100000 = 0.001678

ウェア/ハウス 0.000052 0.001678 一般

0.000758 * 0.000019 * 100000 = 0.001441

0.000115 * 0.001284 * 100000 = 0.014820

グループ/ウェア 0.001441 0.014820 一般

0.000303 * 0.000013 * 0.003166 * 10000000 = 0.000128

0.000054 * 0.000003 * 0.000019 * 10000000 = 0.000000

グループ/販売/会社 0.000128 0.000000 専門

0.004398 * 0.000013 * 0.008593 * 10000000 = 0.005052

0.000038 * 0.000016 * 0.000019 * 10000000 = 0.000000

システム/刷新/計画 0.005052 0.000000 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000057 * 0.002385 * 100000 = 0.013697

システム/所有 0.000001 0.013697 一般

0.000001 * 0.001357 * 100000 = 0.000077
0.000057 * 0.000019 * 100000 = 0.000109
システム/部門 0.000077 0.000109 一般

0.000001 * 0.013569 * 100000 = 0.000773
0.000089 * 0.003046 * 100000 = 0.027018
データ/交換 0.000773 0.027018 一般

0.000101 * 0.000452 * 100000 = 0.004557
0.000018 * 0.000019 * 100000 = 0.000033
デスク/トップ 0.004557 0.000033 専門

0.000001 * 0.000452 * 100000 = 0.000026
0.000099 * 0.002202 * 100000 = 0.021785
パソコン/運用 0.000026 0.021785 一般

0.003033 * 0.024423 * 100000 = 7.408283
0.000005 * 0.000587 * 100000 = 0.000300
ファイル/変換 7.408283 0.000300 専門

0.000051 * 0.000019 * 100000 = 0.000096
0.000001 * 0.000917 * 100000 = 0.000104
ヘルプ/デスク 0.000096 0.000104 一般

0.000001 * 0.000003 * 0.000003 * 0.017187 * 1000000000 = 0.000000
0.000001 * 0.000022 * 0.000003 * 0.000019 * 1000000000 = 0.000000
ヘルプ/デスク/担当/者 0.000000 0.000000 専門

0.000001 * 0.001357 * 100000 = 0.000077
0.000023 * 0.000019 * 100000 = 0.000043
マーケティング/部門 0.000077 0.000043 専門

0.000001 * 0.005638 * 0.000905 * 10000000 = 0.000044
0.000001 * 0.000003 * 0.001725 * 10000000 = 0.000000
モバイル/用/ノート 0.000044 0.000000 専門

$$0.000001 * 0.000019 * 100000 = 0.000001$$

$$0.000035 * 0.000019 * 100000 = 0.000066$$

ユーザー/企業 0.000001 0.000066 一般

$$0.000001 * 0.000452 * 100000 = 0.000026$$

$$0.000032 * 0.000019 * 100000 = 0.000062$$

リース/契約 0.000026 0.000062 一般

$$0.000001 * 0.000452 * 100000 = 0.000026$$

$$0.000032 * 0.001138 * 100000 = 0.003688$$

リース/切れ 0.000026 0.003688 一般

$$0.000001 * 0.000013 * 0.000019 * 10000000 = 0.000000$$

$$0.000085 * 0.000011 * 0.000019 * 10000000 = 0.000000$$

リ/ソース/不足 0.000000 0.000000 一般

$$0.000001 * 0.003618 * 100000 = 0.000206$$

$$0.000001 * 0.000019 * 100000 = 0.000001$$

リプレース/作業 0.000206 0.000001 専門

$$0.000001 * 0.001809 * 100000 = 0.000103$$

$$0.000101 * 0.000019 * 100000 = 0.000191$$

運用/コスト 0.000103 0.000191 一般

$$0.000050 * 0.000235 * 0.000019 * 10000000 = 0.000002$$

$$0.000050 * 0.000180 * 0.000019 * 10000000 = 0.000002$$

運用/コスト/削減 0.000002 0.000002 専門

$$0.000050 * 0.005638 * 0.004523 * 10000000 = 0.012840$$

$$0.000050 * 0.000003 * 0.000220 * 10000000 = 0.000000$$

運用/管理/ツール 0.012840 0.000000 専門

$$0.000001 * 0.000452 * 100000 = 0.000026$$

$$0.000001 * 0.000019 * 100000 = 0.000001$$

営業/所 0.000026 0.000001 専門

$$0.000001 * 0.000019 * 100000 = 0.000001$$

0.000001 * 0.000019 * 100000 = 0.000001

営業/部隊 0.000001 0.000001 専門

0.000001 * 0.000013 * 0.000019 * 10000000 = 0.000000

0.000031 * 0.000322 * 0.000019 * 10000000 = 0.000002

卸売/業者/向け 0.000000 0.000002 一般

0.000253 * 0.022162 * 100000 = 0.560520

0.000031 * 0.000110 * 100000 = 0.000338

会計/モジュール 0.560520 0.000338 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.000019 * 100000 = 0.000001

各/企業 0.000001 0.000001 専門

0.000506 * 0.000013 * 0.000452 * 10000000 = 0.000031

0.000224 * 0.000003 * 0.000019 * 10000000 = 0.000000

管理/ツール/導入 0.000031 0.000000 専門

0.000404 * 0.017187 * 100000 = 0.695116

0.000001 * 0.000019 * 100000 = 0.000001

関係/者 0.695116 0.000001 専門

0.000354 * 0.005427 * 100000 = 0.191978

0.000001 * 0.000367 * 100000 = 0.000021

技術/サポート 0.191978 0.000021 専門

0.000001 * 0.001809 * 100000 = 0.000103

0.000097 * 0.000073 * 100000 = 0.000706

業務/アプリケーション 0.000103 0.000706 一般

0.000303 * 0.001809 * 100000 = 0.054821

0.000026 * 0.000019 * 100000 = 0.000049

検索/ソフト 0.054821 0.000049 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000028 * 0.000550 * 100000 = 0.001532

原因/究明 0.000001 0.001532 一般

$0.001365 * 0.001644 * 0.023971 * 10000000 = 0.538059$

$0.000016 * 0.000164 * 0.000019 * 10000000 = 0.000001$

故障/発生/率 0.538059 0.000001 専門

$0.000001 * 0.000019 * 100000 = 0.000001$

$0.000028 * 0.002679 * 100000 = 0.007464$

購買/窓口 0.000001 0.007464 一般

$0.000202 * 0.000019 * 100000 = 0.000384$

$0.000001 * 0.000019 * 100000 = 0.000001$

高/スバック 0.000384 0.000001 専門

$0.000001 * 0.000019 * 100000 = 0.000001$

$0.000020 * 0.000019 * 100000 = 0.000038$

三つ/目 0.000001 0.000038 一般

$0.000001 * 0.000019 * 100000 = 0.000001$

$0.000001 * 0.000019 * 100000 = 0.000001$

参加/企業 0.000001 0.000001 専門

$0.000001 * 0.000019 * 100000 = 0.000001$

$0.000026 * 0.003046 * 100000 = 0.007794$

部長/代理 0.000001 0.007794 一般

$0.000001 * 0.036183 * 100000 = 0.002061$

$0.000063 * 0.000019 * 100000 = 0.000120$

資産/管理 0.002061 0.000120 専門

$0.000001 * 0.000452 * 100000 = 0.000026$

$0.000064 * 0.000019 * 100000 = 0.000122$

次期/版 0.000026 0.000122 一般

$0.000001 * 0.003994 * 0.000003 * 0.000019 * 1000000000 = 0.000000$

$0.000001 * 0.000001 * 0.000027 * 0.000019 * 1000000000 = 0.000000$

自動車/用/タイヤ/大手 0.000000 0.000000 専門

$$0.000001 * 0.039349 * 100000 = 0.002241$$

$$0.000045 * 0.000019 * 100000 = 0.000086$$

実用/的 0.002241 0.000086 専門

$$0.000001 * 0.003166 * 100000 = 0.000180$$

$$0.000115 * 0.000073 * 100000 = 0.000843$$

社内/LAN 0.000180 0.000843 一般

$$0.000001 * 0.000019 * 100000 = 0.000001$$

$$0.000115 * 0.000019 * 100000 = 0.000219$$

社内/システム 0.000001 0.000219 一般

$$0.000001 * 0.000019 * 100000 = 0.000001$$

$$0.000115 * 0.000019 * 100000 = 0.000219$$

社内/同士 0.000001 0.000219 一般

$$0.000001 * 0.001527 * 0.001809 * 10000000 = 0.000024$$

$$0.000082 * 0.000115 * 0.000019 * 10000000 = 0.000002$$

社内/標準/ソフト 0.000024 0.000002 専門

$$0.001314 * 0.021257 * 100000 = 2.793499$$

$$0.000083 * 0.000550 * 100000 = 0.004566$$

主/記憶 2.793499 0.004566 専門

$$0.000809 * 0.000470 * 0.000019 * 10000000 = 0.000072$$

$$0.000059 * 0.000038 * 0.000019 * 10000000 = 0.000000$$

処理/性能/不足 0.000072 0.000000 専門

$$0.000556 * 0.004523 * 100000 = 0.251464$$

$$0.000027 * 0.002018 * 100000 = 0.005393$$

障害/監視 0.251464 0.005393 専門

$$0.002022 * 0.004581 * 0.001357 * 10000000 = 0.125714$$

$$0.000001 * 0.000273 * 0.000019 * 10000000 = 0.000000$$

情報/システム/部門 0.125714 0.000000 専門

0.002578 * 0.000019 * 100000 = 0.004899
0.000001 * 0.000019 * 100000 = 0.000001
情報/化 0.004899 0.000001 専門

0.000051 * 0.000019 * 100000 = 0.000096
0.000001 * 0.000019 * 100000 = 0.000001
新/システム 0.000096 0.000001 専門

0.000051 * 0.000452 * 100000 = 0.002292
0.000001 * 0.001872 * 100000 = 0.000106
新/パソコン 0.002292 0.000106 専門

0.000809 * 0.000452 * 100000 = 0.036562
0.000027 * 0.000019 * 100000 = 0.000051
性能/向上 0.036562 0.000051 専門

0.000809 * 0.000019 * 100000 = 0.001537
0.000027 * 0.000019 * 100000 = 0.000051
性能/不足 0.001537 0.000051 専門

0.000051 * 0.022162 * 100000 = 0.112356
0.000005 * 0.000110 * 100000 = 0.000056
生産管理/モジュール 0.112356 0.000056 専門

0.000556 * 0.000019 * 100000 = 0.001056
0.000001 * 0.000019 * 100000 = 0.000001
生産/性 0.001056 0.000001 専門

0.000354 * 0.008593 * 100000 = 0.303973
0.000076 * 0.000019 * 100000 = 0.000145
製品/計画 0.303973 0.000145 専門

0.000354 * 0.002714 * 100000 = 0.096006
0.000076 * 0.000440 * 100000 = 0.003352
製品/品質 0.096006 0.003352 専門

0.000051 * 0.002261 * 100000 = 0.011463

0.000007 * 0.000019 * 100000 = 0.000013
全社/展開 0.011463 0.000013 専門

0.001062 * 0.002714 * 100000 = 0.288174
0.000028 * 0.000019 * 100000 = 0.000053
操作/方法 0.288174 0.000053 専門

0.000051 * 0.001809 * 100000 = 0.009171
0.000001 * 0.000019 * 100000 = 0.000001
総/コスト 0.009171 0.000001 専門

0.000051 * 0.001357 * 100000 = 0.006880
0.000034 * 0.000019 * 100000 = 0.000064
測定/調査 0.006880 0.000064 専門

0.000202 * 0.000019 * 100000 = 0.000384
0.000001 * 0.000019 * 100000 = 0.000001
大/規模 0.000384 0.000001 専門

0.000202 * 0.000019 * 100000 = 0.000384
0.000001 * 0.000019 * 100000 = 0.000001
大量/リプレース 0.000384 0.000001 専門

0.000202 * 0.000452 * 100000 = 0.009140
0.000001 * 0.000019 * 100000 = 0.000001
大量/導入 0.009140 0.000001 専門

0.001011 * 0.000019 * 100000 = 0.001921
0.000004 * 0.000019 * 100000 = 0.000008
値/自体 0.001921 0.000008 専門

0.000050 * 0.000013 * 0.000019 * 10000000 = 0.000000
0.000001 * 0.000273 * 0.000220 * 10000000 = 0.000001
定期/ご/購読 0.000000 0.000001 一般

0.000051 * 0.000019 * 100000 = 0.000096
0.000007 * 0.000550 * 100000 = 0.000375

添付/ファイル 0.000096 0.000375 一般

0.001820 * 0.001809 * 100000 = 0.329237

0.000105 * 0.000220 * 100000 = 0.002302

電子/メール 0.329237 0.002302 専門

0.000253 * 0.003166 * 100000 = 0.080074

0.000070 * 0.000019 * 100000 = 0.000133

電力/会社 0.080074 0.000133 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000038 * 0.000294 * 100000 = 0.001103

東洋/ゴム 0.000001 0.001103 一般

0.000556 * 0.004581 * 0.000019 * 10000000 = 0.000484

0.000080 * 0.000273 * 0.000019 * 10000000 = 0.000004

統合/システム/部長 0.000484 0.000004 専門

0.000556 * 0.000705 * 0.010403 * 10000000 = 0.040779

0.000080 * 0.000148 * 0.000550 * 10000000 = 0.000065

統合/業務/パッケージ 0.040779 0.000065 専門

0.000607 * 0.013116 * 100000 = 0.795701

0.000005 * 0.000019 * 100000 = 0.000010

動作/環境 0.795701 0.000010 専門

0.000001 * 0.025780 * 100000 = 0.001469

0.000011 * 0.000073 * 100000 = 0.000079

日経/コンピュータ 0.001469 0.000079 専門

0.000101 * 0.000019 * 100000 = 0.000192

0.000001 * 0.000037 * 100000 = 0.000002

日本/マクドナルド 0.000192 0.000002 専門

0.000051 * 0.010855 * 100000 = 0.055032

0.000001 * 0.000019 * 100000 = 0.000001

販売/戦略 0.055032 0.000001 専門

0.000505 * 0.000019 * 100000 = 0.000960

0.000098 * 0.002752 * 100000 = 0.026914

副/主任 0.000960 0.026914 一般

0.001567 * 0.000452 * 100000 = 0.070832

0.000076 * 0.001872 * 100000 = 0.014263

標準/パソコン 0.070832 0.014263 専門

0.000101 * 0.005991 * 0.001809 * 10000000 = 0.010960

0.000026 * 0.000076 * 0.000019 * 10000000 = 0.000000

表/計算/ソフト 0.010960 0.000000 専門

0.000253 * 0.000019 * 100000 = 0.000481

0.000048 * 0.000019 * 100000 = 0.000092

部品/代 0.000481 0.000092 専門

0.000101 * 0.000905 * 100000 = 0.009125

0.000018 * 0.000844 * 100000 = 0.001488

復旧/処置 0.009125 0.001488 専門

0.000253 * 0.001809 * 100000 = 0.045753

0.000048 * 0.000019 * 100000 = 0.000092

分析/ソフト 0.045753 0.000092 専門

0.000960 * 0.000019 * 100000 = 0.001825

0.000052 * 0.000550 * 100000 = 0.002877

文書/ファイル 0.001825 0.002877 一般

0.000001 * 0.039349 * 100000 = 0.002241

0.000001 * 0.000019 * 100000 = 0.000001

本格/的 0.002241 0.000001 専門

0.000001 * 0.000452 * 100000 = 0.000026

0.000001 * 0.000019 * 100000 = 0.000001

本格/導入 0.000026 0.000001 専門

$$0.000051 * 0.000019 * 100000 = 0.000096$$

$$0.000001 * 0.000019 * 100000 = 0.000001$$

本/記事 0.000096 0.000001 専門

$$0.000001 * 0.000019 * 100000 = 0.000001$$

$$0.000023 * 0.000019 * 100000 = 0.000043$$

無断/転載 0.000001 0.000043 一般

$$0.000001 * 0.017187 * 100000 = 0.000979$$

$$0.000098 * 0.000019 * 100000 = 0.000186$$

利用/者 0.000979 0.000186 専門

$$0.000001 * 0.001357 * 100000 = 0.000077$$

$$0.000098 * 0.000019 * 100000 = 0.000186$$

利用/部門 0.000077 0.000186 一般

$$0.000001 * 0.000013 * 0.000019 * 10000000 = 0.000000$$

$$0.000014 * 0.000003 * 0.000019 * 10000000 = 0.000000$$

量販/店/向け 0.000000 0.000000 一般

文書 2

本文

インターネット・ビジネスの行方占う“試金石”

オンライン証券が続々登場

10月1日の株式委託売買手数料の自由化を目前に控え、インターネットを利用するオンライン証券取引サービスが続々と登場している。米国で実績を持つイー・トレード証券など外資系だけでなく、ソニーなどの異業種からの参入も多い。日興証券など大手の証券会社もオンライン専業の子会社を立ち上げるなど、インターネット・ビジネスを巡る動きが活発になってきた。オンライン証券各社は取引手数料の安さを“武器”にビジネスの立ち上げを図る。

関 信浩

本記事は日経コンピュータ99年9月13日号からの抜粋です。そのため図や表が一部割愛されていることをあらかじめご了承ください。なお本号のご購入はバックナンバー、または日経コンピュータの定期ご購読をご利用ください。

「ネットワーク時代の新しいベンチャ企業に投資するのは大きなチャレンジだ。成功するように全力でバックアップする」一。ソニーの出井伸之社長は8月25日、同社が主要株主として設立した「オンライン」証券会社、マネックス証券（東京都千代田区）の記者会見で力説した（写真1）。

マネックス証券は米国屈指の投資銀行、ゴールドマン・サックス証券出身の松本大氏がソニーと共同で今年4月に設立した。この10月からインターネットと電話による証券取引サービスを始めるが、目玉は顧客がインターネットを使って取引の注文を出す「インターネット取引」である。松本社長は「単に証券取引の場をインターネットに移すだけでなく、ネットワーク時代を強く意識した、新しい金融サービスを提供する」と意気込む。

マネックス証券の共同出資者にはソニーのほか、リクルート（東京都中央区）やインターネットイ

ニシアティブ（東京都千代田区）など、インターネット・ビジネスを強く推進していることで著名な企業が名を連ねる。

9月1日には、インターネット上の金融事業に特化したソフトバンクの子会社、ソフトバンク・ファイナンス（東京都千代田区）が「インターネット金融戦略」を発表。ソフトバンク傘下のオンライン証券であるイー・トレード証券（東京都千代田区）などのさまざまな金融サービスを、インターネットで統合して「WWW上の“ワンストップ金融サービス”を実現し、世界初のインターネット総合金融サービス会社になる」（北尾吉孝社長）と宣言した。

外資、異業種、既存組が入り乱れる

オンライン証券では、証券取引の受け付けをインターネットや電話に絞り込むことによって、コストの大幅な削減を目指す。証券取引はモノの流れを伴わないため、「インターネットに最も適したビジネスの一つ」とみられている。「米国では証券取引の30%がインターネット取引に移行している」（日興証券の金子昌資社長）。すでに国内でもオンライン取引に比重を移している松井証券では、「すでにオンライン取引による手数料収入が全体の5割を超えた」（佐藤義仁取締役）。

10月1日の株式委託売買手数料の自由化を目前に控えて、外資系の進出や異業種からの参入が相次いでいる。ソニー以外にも、この4月に旅行代理店のエイチ・アイ・エスが協立証券に出資して設立したエイチ・アイ・エス協立証券をはじめとして、富士通などが2000年1月に設立するインターネット・トレーディング証券（仮称）など目白押しだ（表1）。

ベンチャー企業も登場した。インターネット上の株式関連フォーラムで著名な臼田琢美氏などが中心になって7月に設立した日本オンライン証券（東京都千代田区）である。伊藤忠商事や第一勧業銀行、マイクロソフトなどの出資を受けている。

米国でオンライン証券最大手のチャールズ・シュワブも、7月に東京海上火災保険と合併でシュワブ東京海上証券（東京都千代田区）を設立。この秋から営業を開始する。米DLJダイレクト証券も、住友銀行など住友グループや大和証券の出資を受け、オンライン専門の「ディーエルジェイダイレクト・エスエフジー証券（東京都千代田区）」を設立し、6月にサービスを開始した。

続きは日経コンピュータ99年9月13日号をお読み下さい。この号のご購入はバックナンバー、または日経コンピュータの定期ご購入をご利用ください。

このホームページに掲載されている記事・写真・図表などの無断転載を禁じます。

判別結果

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.000019 * 100000 = 0.000001

ご/購入 0.000001 0.000001 専門

0.000001 * 0.003166 * 100000 = 0.000180

0.000001 * 0.000019 * 100000 = 0.000001

ご/利用 0.000180 0.000001 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.000477 * 100000 = 0.000027

ご/了承 0.000001 0.000027 一般

0.000001 * 0.000013 * 0.010855 * 10000000 = 0.000001

0.000009 * 0.000443 * 0.000019 * 10000000 = 0.000001

インターネット/金融/戦略 0.000001 0.000001 専門

0.000253 * 0.000019 * 100000 = 0.000481

0.000003 * 0.000019 * 100000 = 0.000005

インターネット/取引 0.000481 0.000005 専門

0.000001 * 0.000003 * 0.000117 * 0.001058 * 0.000000 * 1000000000 = 0.000000

0.000009 * 0.000284 * 0.000202 * 0.000382 * 0.000000 * 1000000000 = 0.000000

インターネット/総合/金融/サービス/会社 0.000000 0.000000 一般

0.000001 * 0.000003 * 0.000003 * 0.000019 * 1000000000 = 0.000000

0.000001 * 0.000044 * 0.000003 * 0.000019 * 1000000000 = 0.000000

エイチ・アイ・エス/協/立証/券 0.000000 0.000000 一般

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.001762 * 100000 = 0.000100

エスエフジー/証券 0.000001 0.000100 一般

0.001163 * 0.000019 * 100000 = 0.002209

0.000026 * 0.000019 * 100000 = 0.000049

オンライン/取引 0.002209 0.000049 専門

0.001163 * 0.000019 * 100000 = 0.002209

0.000026 * 0.001762 * 100000 = 0.004508

オンライン/証券 0.002209 0.004508 一般

0.000758 * 0.000013 * 0.000019 * 10000000 = 0.000002

0.000012 * 0.000322 * 0.000019 * 10000000 = 0.000001

オンライン/証券/各社 0.000002 0.000001 専門

0.000758 * 0.000003 * 0.000353 * 0.001357 * 1000000000 = 0.000001

0.000012 * 0.000076 * 0.000003 * 0.000019 * 1000000000 = 0.000000

オンライン/証券/最大/手 0.000001 0.000000 専門

0.000758 * 0.000013 * 0.028494 * 10000000 = 0.002888

0.000012 * 0.000011 * 0.000019 * 10000000 = 0.000000

オンライン/証券取引/サービス 0.002888 0.000000 専門

0.001163 * 0.000019 * 100000 = 0.002209

0.000026 * 0.000514 * 100000 = 0.001315

オンライン/専業 0.002209 0.001315 専門

0.000001 * 0.000905 * 100000 = 0.000052

0.000027 * 0.000019 * 100000 = 0.000051

ゴールド/マン 0.000052 0.000051 専門

0.000001 * 0.000013 * 0.000019 * 10000000 = 0.000000

0.000002 * 0.000322 * 0.000019 * 10000000 = 0.000000

サックス/証券/出身 0.000000 0.000000 一般

0.000001 * 0.000003 * 0.000003 * 0.000019 * 1000000000 = 0.000000

0.000005 * 0.000377 * 0.000038 * 0.001762 * 1000000000 = 0.000000

シュワブ/東京/海上/証券 0.000000 0.000000 一般

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.001174 * 100000 = 0.000067

ソフトバンク/傘下 0.000001 0.000067 一般

0.000001 * 0.000019 * 100000 = 0.000001
0.000001 * 0.001762 * 100000 = 0.000100
ダイレクト/証券 0.000001 0.000100 一般

0.000001 * 0.000019 * 100000 = 0.000001
0.000001 * 0.001762 * 100000 = 0.000100
トレーディング/証券 0.000001 0.000100 一般

0.000001 * 0.000019 * 100000 = 0.000001
0.000004 * 0.001762 * 100000 = 0.000701
トレード/証券 0.000001 0.000701 一般

0.001264 * 0.000019 * 100000 = 0.002402
0.000047 * 0.000019 * 100000 = 0.000089
ネットワーク/時代 0.002402 0.000089 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000001 * 0.000019 * 100000 = 0.000001
ベンチャ/企業 0.000001 0.000001 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000001 * 0.001762 * 100000 = 0.000100
マネックス/証券 0.000001 0.000100 一般

0.000050 * 0.000003 * 0.000003 * 0.028494 * 1000000000 = 0.000000
0.000035 * 0.000005 * 0.000443 * 0.000019 * 1000000000 = 0.000000
ワン/ストップ/金融/サービス 0.000000 0.000000 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000034 * 0.001541 * 100000 = 0.005170
異/業種 0.000001 0.005170 一般

0.000001 * 0.019900 * 100000 = 0.001134
0.000015 * 0.000019 * 100000 = 0.000028
外資/系 0.001134 0.000028 専門

0.000050 * 0.000118 * 0.000019 * 10000000 = 0.000001

0.000001 * 0.000003 * 0.000019 * 10000000 = 0.000000
株式/関連/フォーラム 0.000001 0.000000 専門

0.000001 * 0.000452 * 100000 = 0.000026
0.000049 * 0.000019 * 100000 = 0.000094
既存/組 0.000026 0.000094 一般

0.000001 * 0.000019 * 100000 = 0.000001
0.000056 * 0.000019 * 100000 = 0.000107
記者/会見 0.000001 0.000107 一般

0.000050 * 0.000013 * 0.017187 * 10000000 = 0.000116
0.000001 * 0.000071 * 0.000019 * 10000000 = 0.000000
共同/出資/者 0.000116 0.000000 専門

0.000001 * 0.000013 * 0.000019 * 10000000 = 0.000000
0.000016 * 0.000003 * 0.000019 * 10000000 = 0.000000
協/立証/券 0.000000 0.000000 一般

0.000051 * 0.028494 * 100000 = 0.144458
0.000113 * 0.000019 * 100000 = 0.000214
金融/サービス 0.144458 0.000214 専門

0.000051 * 0.000019 * 100000 = 0.000096
0.000113 * 0.000019 * 100000 = 0.000214
金融/事業 0.000096 0.000214 一般

0.000505 * 0.000019 * 100000 = 0.000960
0.000001 * 0.000019 * 100000 = 0.000001
自由/化 0.000960 0.000001 専門

0.000152 * 0.000019 * 100000 = 0.000288
0.000001 * 0.000917 * 100000 = 0.000052
主要/株主 0.000288 0.000052 専門

0.000001 * 0.012212 * 100000 = 0.000696
0.000001 * 0.000019 * 100000 = 0.000001

住友/グループ 0.000696 0.000001 専門

0.000001 * 0.003166 * 100000 = 0.000180

0.000061 * 0.000019 * 100000 = 0.000117

証券/会社 0.000180 0.000117 専門

0.000001 * 0.028494 * 100000 = 0.001623

0.000005 * 0.000019 * 100000 = 0.000010

証券取引/サービス 0.001623 0.000010 専門

0.000101 * 0.000019 * 100000 = 0.000192

0.000001 * 0.000019 * 100000 = 0.000001

世界/初 0.000192 0.000001 専門

0.000050 * 0.000013 * 0.000019 * 10000000 = 0.000000

0.000001 * 0.000273 * 0.000220 * 10000000 = 0.000001

定期/ご/購読 0.000000 0.000001 一般

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.000019 * 100000 = 0.000001

投資/銀行 0.000001 0.000001 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000016 * 0.000019 * 100000 = 0.000030

特/化 0.000001 0.000030 一般

0.000001 * 0.025780 * 100000 = 0.001469

0.000011 * 0.000073 * 100000 = 0.000079

日経/コンピュータ 0.001469 0.000079 専門

0.000961 * 0.000118 * 0.000019 * 10000000 = 0.000021

0.000001 * 0.000005 * 0.001762 * 10000000 = 0.000000

日本/オンライン/証券 0.000021 0.000000 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.001284 * 100000 = 0.000073

米国/屈指 0.000001 0.000073 一般

$$0.000051 * 0.000019 * 100000 = 0.000096$$

$$0.000001 * 0.000019 * 100000 = 0.000001$$

本/記事 0.000096 0.000001 専門

$$0.000001 * 0.000019 * 100000 = 0.000001$$

$$0.000023 * 0.000019 * 100000 = 0.000043$$

無断/転載 0.000001 0.000043 一般

$$0.000001 * 0.000019 * 100000 = 0.000001$$

$$0.000012 * 0.000917 * 100000 = 0.001095$$

立ち/上げ 0.000001 0.001095 一般

$$0.000001 * 0.000013 * 0.000452 * 10000000 = 0.000000$$

$$0.000127 * 0.000169 * 0.000019 * 10000000 = 0.000004$$

旅行/代理/店 0.000000 0.000004 一般

文書 3

本文

異なるシステムを統合するための“切り札”

E A I ツールが続々登場

異なるシステム間を連携させるためのミドルウェアである「E A I（エンタープライズ・アプリケーション・インテグレーション）ツール」が注目を集めている。企業内のシステム統合や企業間システム連携などのニーズを受けて、新製品が相次いでいる。異なるシステム同士を接続する「接続アダプタ」や、互いのシステムのデータ項目を対応づける「マッピング機能」を備える。プラットフォームの違いを意識せず複数システム間でデータを連携できる。

高下 義弘

本記事は日経コンピュータ99年9月13日号からの抜粋です。そのため図や表が一部割愛されていることをあらかじめご了承ください。なお本号のご購入はバックナンバー、または日経コンピュータの定期ご購読をご利用ください。

異なるシステムを連携させるためのミドルウェア製品「E A I（エンタープライズ・アプリケーション・インテグレーション）ツール」が続々と登場している。E A I ツールは、複数の異なるシステム間でデータをやり取りする際に必要になる、通信機能やインタフェース機能、データ変換機能などを備える。すでに国内でシステムの稼働実績がある「Mercator」（三井物産が販売）をはじめとして、全世界で1000社の実績がある「Data Gate」の新版「E* Gate」（販売はソフトウェア・テクノロジー・コーポレーション＝STC、東京都港区＝）が9月に発売されるなど、来年の前半までに10製品弱のE A I ツールが市場に出そろう（35ページの表1）。

連携のニーズは増加している

E A I ツールが相次いで登場している背景には、インターネットやERPパッケージ（統合業務パッケージ）の普及によって、企業の内外を問わずにシステム同士を連携させるニーズが高まっている

ことがある。

連携するシステムが少ないうちはよいが、連携させるシステムが増えると、システムごとに連携用プログラムを開発して、保守するのは膨大な手間になる。EAIツールはこうしたシステム間連携において、いわば“ハブ”の役割を果たすミドルウェアである（図1）。

「EAIツールを使えば、システムごとに連携プログラムをいちいち記述する手間を省ける。新たに連携させるシステムを追加する際にも、ほかのシステムにほとんど手を入れる必要がない」（三井物産ITマーケティング事業部アドバンスマーケティング事業室ダイレクトマーケティングチームの齋藤良樹チームリーダー）。

EAIツールの適用例としては、ERPパッケージ（統合業務パッケージ）と既存システムの連携、サプライチェーン管理を実現するためのシステム連携、企業の合併によるシステム統合などがある。「最近、特にニーズが多いのがERPパッケージと既存システムの連携」（NTT東日本でERPパッケージを使ったシステム・インテグレーションに携わっている尾崎光則法人営業本部システムサービス部ソリューショングループERP担当課長）である。

ERPパッケージは本来、企業の基幹系システムをすべて置き換えることを目指しているが、現実にそうしているユーザー企業はあまりない。会計システムだけをERPパッケージの会計モジュールで置き換えて、生産管理システムや販売管理システムは既存のメインフレームで稼働を継続させるなど、部分的に導入している企業がほとんどである。

ERPパッケージと既存システムを一緒に使っている場合には、互いのシステム間の連携が問題になる。連携させるシステムの種類が増えてくると、連携の仕組みを開発する手間が膨大になってくる。「ERPパッケージを導入する際に、こうした連携の仕組みを開発する手間は無視できないほど大きい」（NTT東日本の尾崎担当課長）。

サプライチェーン管理のニーズも今後ますます増加しそうだ。「ビジネスにおける競争が激しくなり、経営環境の変化のスピードが速くなるにつれて、取引先をより条件の良い企業に変えることが頻繁に起こり得る。この際には、いかに素早く、かつコストを抑えて相手先のシステムと連携させられるかが競争に勝つためのポイントになる」（電通国際情報サービス＝東京都中野区＝の山川裕EC事業部ソリューションコンサルティング部部長）。EAIツールはこうした問題を解決するための“切り札”になり得る。

続きは日経コンピュータ99年9月13日号をお読み下さい。この号のご購入はバックナンバー、または日経コンピュータの定期ご購読をご利用ください。

このホームページに掲載されている記事・写真・図表などの無断転載を禁じます。

判別結果

0.000001 * 0.000003 * 0.000003 * 0.010403 * 1000000000 = 0.000000
0.000001 * 0.000001 * 0.000005 * 0.000550 * 1000000000 = 0.000000
E/R/P/パッケージ 0.000000 0.000000 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000070 * 0.000037 * 100000 = 0.000259
N T T/東日本 0.000001 0.000259 一般

0.000001 * 0.000019 * 100000 = 0.000001
0.000001 * 0.000019 * 100000 = 0.000001
ご/購入 0.000001 0.000001 専門

0.000001 * 0.003166 * 100000 = 0.000180
0.000001 * 0.000019 * 100000 = 0.000001
ご/利用 0.000180 0.000001 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000001 * 0.000477 * 100000 = 0.000027
ご/了承 0.000001 0.000027 一般

0.000960 * 0.000019 * 100000 = 0.001825
0.000001 * 0.000019 * 100000 = 0.000002
インタフェース/機能 0.001825 0.000002 専門

0.000001 * 0.000235 * 0.036183 * 10000000 = 0.000073
0.000001 * 0.000093 * 0.000019 * 10000000 = 0.000000
サプライ/チェーン/管理 0.000073 0.000000 専門

0.000001 * 0.000452 * 100000 = 0.000026
0.000057 * 0.000019 * 100000 = 0.000109
システム/ごと 0.000026 0.000109 一般

0.000001 * 0.001809 * 100000 = 0.000103
0.000057 * 0.000019 * 100000 = 0.000109
システム/間 0.000103 0.000109 一般

0.004398 * 0.002349 * 0.000019 * 10000000 = 0.001963

0.000038 * 0.000003 * 0.000881 * 10000000 = 0.000001

システム/間/連携 0.001963 0.000001 専門

0.000001 * 0.001809 * 100000 = 0.000103

0.000057 * 0.001798 * 100000 = 0.010326

システム/統合 0.000103 0.010326 一般

0.000001 * 0.000019 * 100000 = 0.000001

0.000057 * 0.000019 * 100000 = 0.000109

システム/同士 0.000001 0.000109 一般

0.000001 * 0.000019 * 100000 = 0.000001

0.000057 * 0.000881 * 100000 = 0.005059

システム/連携 0.000001 0.005059 一般

0.000001 * 0.030303 * 100000 = 0.001726

0.000089 * 0.003156 * 100000 = 0.027994

データ/項目 0.001726 0.027994 一般

0.000001 * 0.002702 * 0.000019 * 10000000 = 0.000000

0.000057 * 0.000038 * 0.000019 * 10000000 = 0.000000

データ/変換/機能 0.000000 0.000000 専門

0.000354 * 0.000019 * 100000 = 0.000672

0.000001 * 0.000019 * 100000 = 0.000001

マッピング/機能 0.000672 0.000001 専門

0.000051 * 0.000019 * 100000 = 0.000096

0.000011 * 0.001284 * 100000 = 0.001387

ミドル/ウェア 0.000096 0.001387 一般

0.000001 * 0.000013 * 0.000019 * 10000000 = 0.000000

0.000014 * 0.000049 * 0.000019 * 10000000 = 0.000000

ミドル/ウェア/製品 0.000000 0.000000 一般

0.000001 * 0.005427 * 100000 = 0.000309

0.000001 * 0.000294 * 100000 = 0.000033

メイン/フレーム 0.000309 0.000033 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000035 * 0.000019 * 100000 = 0.000066

ユーザー/企業 0.000001 0.000066 一般

0.000152 * 0.000019 * 100000 = 0.000288

0.000023 * 0.000019 * 100000 = 0.000043

稼働/実績 0.000288 0.000043 専門

0.000253 * 0.000019 * 100000 = 0.000481

0.000031 * 0.000019 * 100000 = 0.000058

会計/システム 0.000481 0.000058 専門

0.000253 * 0.022162 * 100000 = 0.560520

0.000031 * 0.000110 * 100000 = 0.000338

会計/モジュール 0.560520 0.000338 専門

0.000050 * 0.000705 * 0.004581 * 0.000019 * 1000000000 = 0.000003

0.000001 * 0.000191 * 0.000273 * 0.000881 * 1000000000 = 0.000000

企業/間/システム/連携 0.000003 0.000000 専門

0.000051 * 0.001357 * 100000 = 0.006880

0.000001 * 0.000019 * 100000 = 0.000001

企業/内 0.006880 0.000001 専門

0.000001 * 0.000353 * 0.000019 * 10000000 = 0.000000

0.000014 * 0.000003 * 0.000019 * 10000000 = 0.000000

基幹/系/システム 0.000000 0.000000 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000049 * 0.000019 * 100000 = 0.000094

既存/システム 0.000001 0.000094 一般

0.000354 * 0.013116 * 100000 = 0.463972

0.000001 * 0.000019 * 100000 = 0.000001
経営/環境 0.463972 0.000001 専門

0.000101 * 0.000905 * 100000 = 0.009125
0.000101 * 0.000019 * 100000 = 0.000191
取引/先 0.009125 0.000191 専門

0.000051 * 0.000019 * 100000 = 0.000096
0.000001 * 0.000019 * 100000 = 0.000001
新/製品 0.000096 0.000001 専門

0.000051 * 0.000019 * 100000 = 0.000096
0.000005 * 0.000019 * 100000 = 0.000010
生産管理/システム 0.000096 0.000010 専門

0.000354 * 0.000019 * 100000 = 0.000672
0.000076 * 0.000367 * 100000 = 0.002796
製品/弱 0.000672 0.002796 一般

0.000556 * 0.020805 * 100000 = 1.156689
0.000030 * 0.000019 * 100000 = 0.000056
接続/アダプタ 1.156689 0.000056 専門

0.000303 * 0.000905 * 100000 = 0.027426
0.000001 * 0.000019 * 100000 = 0.000001
全/世界 0.027426 0.000001 専門

0.000001 * 0.000905 * 100000 = 0.000052
0.000090 * 0.000019 * 100000 = 0.000171
相手/先 0.000052 0.000171 一般

0.000001 * 0.000019 * 100000 = 0.000001
0.000115 * 0.000019 * 100000 = 0.000219
担当/課長 0.000001 0.000219 一般

0.001466 * 0.000019 * 100000 = 0.002786
0.000001 * 0.000019 * 100000 = 0.000001

通信/機能 0.002786 0.000001 専門

0.000050 * 0.000013 * 0.000019 * 1000000 = 0.000000

0.000001 * 0.000273 * 0.000220 * 1000000 = 0.000001

定期/ご/購読 0.000000 0.000001 一般

0.000303 * 0.000019 * 100000 = 0.000576

0.000056 * 0.000019 * 100000 = 0.000107

適用/例 0.000576 0.000107 専門

0.000556 * 0.000705 * 0.010403 * 1000000 = 0.040779

0.000080 * 0.000148 * 0.000550 * 1000000 = 0.000065

統合/業務/パッケージ 0.040779 0.000065 専門

0.000001 * 0.025780 * 100000 = 0.001469

0.000011 * 0.000073 * 100000 = 0.000079

日経/コンピュータ 0.001469 0.000079 専門

0.000151 * 0.005638 * 0.000019 * 1000000 = 0.000162

0.000001 * 0.000003 * 0.000019 * 1000000 = 0.000000

販売/管理/システム 0.000162 0.000000 専門

0.001213 * 0.039349 * 100000 = 4.774328

0.000074 * 0.000019 * 100000 = 0.000140

部分/的 4.774328 0.000140 専門

0.000051 * 0.000019 * 100000 = 0.000096

0.000001 * 0.000019 * 100000 = 0.000001

本/記事 0.000096 0.000001 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000023 * 0.000019 * 100000 = 0.000043

無断/転載 0.000001 0.000043 一般

0.000001 * 0.000019 * 100000 = 0.000001

0.000022 * 0.000019 * 100000 = 0.000041

連携/プログラム 0.000001 0.000041 一般

0.000001 * 0.005638 * 0.000019 * 10000000 = 0.000001

0.000005 * 0.000003 * 0.000019 * 10000000 = 0.000000

連携/用/プログラム 0.000001 0.000000 専門

文書 4

本文

まっくらいふ

インターネットに接続できないのですが？

TCP/IP の設定で、IP アドレスを手入力してみよう！

何が大丈夫で どんな環境で問題が起きるのか？

iBook や iMac、Power Mac G4、それに Mac OS9 のユーザーの一部で、インターネットに接続できないトラブルが発生している。その影響で Mac の起動時間が異常に長くなったり、起動不能になる事態も起きているようだ。

だが、上記のユーザーのすべてに問題が起きるわけではない。インターネットの接続方法によって事態は異なるのだ。一体どういう人に危険性があるのだろうか。

まず、アナログ電話回線とモデムを使って PPP 接続でインターネットを利用している人には、この問題は発生しない（図 1）。もしも接続できないとしたら、それは別のトラブルだろう。

次に、ISDN 回線の TA にモデムポートや USB ポートで、いわゆる「シリアル直結」をしている人も大丈夫だろう。このケースでも通常は PPP 接続のはずだ。

問題となるのは、「DHCP サーバ」と呼ばれる機能を利用している場合だ（図 2）。たとえば、ISDN 回線にダイヤルアップルーターを接続しているケース。これは小規模なオフィスなどに多く、Ethernet の LAN 経由で複数のパソコンがインターネットを利用できるのだが、ダイヤルアップルーター自体が DHCP サーバを兼ねている。たとえパソコンが 1 台でも、ダイヤルアップルーターを使用していればこの分類に含まれると考えていいだろう。それ以外では、ケーブルテレビの回線とケーブルモデムを利用するケースと実際に DHCP サーバのマシンを設置しているケース。

前述の環境で常に問題が発生するのかわかるとは必ずしもそうでもない。使用しているダイヤルアップルーターやケーブルモデムの機種などによって、問題が起きる場合と起きない場合がある。

原因は、Mac OS のなかでネットワーク環境を管理する Open Transport の仕様が、旧来のものと変わったことにある。アップルとしては新仕様に準拠したのだが、現実には古い仕様のネットワーク環境が多く残されていて、それらとの相性が悪くなってしまったのだろう。

で、何が問題かという、インターネットに接続するコンピューターを識別するための「IP アドレス」の発給がスムーズにできなくなったことだ。大企業や研究機関の常時稼働しているサーバなどの IP アドレスは固定だが、我々のような一般ユーザー

がインターネットを利用する場合、この IP アドレスは、空いているものが一時的に貸し出される。この貸し出し作業を行なうもののひとつが DHCP サーバーだ。通常 IP アドレスの貸し出しに際しては制限時間が設定され、制限時間をオーバーする場合には更新の手続きが必要。

とはいえ、これらはユーザー自身がタッチすることではなく、Open Transport と DHCP サーバー間の問題なのである。新しい Open Transport は途方もない制限時間を要求するので、DHCP サーバーによっては妥当な時間に制限する返事を返す。だが、場合によっては要求を拒否することもあり、そうすると拒否された Mac 側は再び同じ要求を出し、またしても拒否ということが延々と続くのである。これをユーザー側から見ると、Mac がフリーズしているように見えるのだ。

また、新しい Open Transport では IP アドレスの更新手続きにも問題がある。制限時間が 2 時間以上になっていると、更新手続きを行なわないのだ。それなのに、時間がオーバーしても前の IP アドレスを使い続けるため、他のパソコンと IP アドレスが重複し、ネットワーク上で問題を起こす可能性も秘めている。

図 1. このように PPP 接続しているユーザーは問題ない

図 2. このように Ethernet 経由で「DHCP サーバを参照」という設定になっていると問題が発生する可能性がある 図 3. 問題が起きるようなら、「設定方法」のポップアップメニューを使って「手入力」に切り替えよう。IP アドレスが他のパソコンと重複しないように設定するのがポイント。

具体的な対応策は？

さて、前述した条件にあてはまるユーザーで、実際にトラブルが発生したらどうすればよいか？

まず、Mac の起動時に問題が起きて、起動不能あるいは起動時間が異常に長いような場合には、Ethernet のケーブルを抜いておこう。Mac が起動したら差し込みなおせばいい。

インターネットへの接続を回復する手段のうち、もっとも安全で確実なのは、PPP 接続に切り替えることだ。だが、これは通信環境の大幅な悪化になるので、一時しのぎにすぎない。また、Mac OS 8.6 以前の環境に戻すという手もあるが、これも消極的な手段だし、新機種では無理である。

次善の策としては、IP アドレスを「手入力」に切り替えて、固定の IP アドレスを使う方法がある。「TCP/IP」コントロールパネルを「手入力」に切り替えたら、自分で IP アドレスを設定しよう(図 3)。多くのダイヤルアップルーターの「ルータアドレス」は「192.168.0.1」になっているが、念のためマニュアルの仕様書などを確認して欲しい。「サブネットマスク」は「255.255.255.0」で、肝心の「IP アドレス」は他のパソコンと重複しないように、末尾の数字を調整しよう。

たまたま「DHCPサーバ」の設定のままで接続できたとしても、それだけでは安心できない。前述した2時間以上の制限時間というトラブルの可能性があるので。可能ならば、DHCPサーバー側で2時間未満に制限するようにしたい。

なお、米国では1月4日にOpen Transport Updater ver. 2.6の配布が開始された。[\(http://asu.info.apple.com/\)](http://asu.info.apple.com/) これにより上記の症状は解決される。アップルジャパンからも日本語OSに対応した同アップデーターが配布されるはずだ。

その他本誌に記載のトラブルと対処法

2000年に電子ブックリーダー“MS Reader”を発売

最初に登壇したMicrosoft社Dick Brass氏の講演テーマは、日本でも注目を集めている電子ブック“Electric Book”である。講演の冒頭では、書籍から電子ブックへの進化をたどるビデオが流された。ここには、2枚の表示液晶で、本来の書籍の形態である見開き表示を可能にする電子ブック用の読書端末も登場した。

キーノートスピーチの最初に登場したマイクロソフト社のDick Brass氏
電子ブックの先駆けは、古く、'81年にRandom House社が開発した電子辞書に遡るが、特に'95年～'97年に掛けて、急ピッチな開発が進んだ。その間に、読書端末を含めた電子ブックの販売価格が、15万ドルから2万5000ドルに下がり、総出荷台数が約1.5倍になった。この分野の先駆者であるRandom House社は、IBM PCで見る形を採用した。しかし、'98年にハンディー端末で読書を可能にするRocket eBookまでに、最も商業的に成功したのは、ソニーのその名も“電子ブック”であった。

これからも電子ブックの開発は進むであろうが、マイクロソフト社は、将来の電子ブックに使われるプラットフォームのスペックは低いもので十分と考えている。この意味では、プラットフォームの準備はすでに終わっていて、周辺技術の詰めを待つ状態になっている。ハードウェアの分野では優れた性能で低コストという条件を満たす液晶の登場、ソフトウェアの分野ではコピープロテクトなどセキュリティーの技術を待つのみである。

マイクロソフト社は、同時に2000年以降の予測を公開した。ここ数年は読書端末にはPCが使われるが、2002年にはいよいよ600dpiの表示解像度をもつ専用読書端末が登場するという。その翌年には、1回の充電での連続表示時間が8時間、価格が89ドルから899ドル(安価な方はモノクロ)に、2006年にはスタンド型の“Kiosk eBook”が登場。マガジンなど読書のコンテンツをそこから個人の端末にダウンロードする時代がくるという。

判別結果

0.000303 * 0.004111 * 0.000019 * 1000000 = 0.000237
0.000001 * 0.000339 * 0.000183 * 1000000 = 0.000001
2/時間/未満 0.000237 0.000001 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000001 * 0.000019 * 100000 = 0.000001
8/時間 0.000001 0.000001 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000001 * 0.000037 * 100000 = 0.000002
DHCP/サーバー 0.000001 0.000002 一般

0.000001 * 0.000013 * 0.001809 * 1000000 = 0.000000
0.000001 * 0.000003 * 0.000019 * 1000000 = 0.000000
DHCP/サーバー/間 0.000000 0.000000 専門

0.000001 * 0.000013 * 0.004071 * 1000000 = 0.000000
0.000001 * 0.000003 * 0.000019 * 1000000 = 0.000000
DHCP/サーバー/側 0.000000 0.000000 専門

0.000001 * 0.037540 * 100000 = 0.002138
0.000001 * 0.000019 * 100000 = 0.000001
IP/アドレス 0.002138 0.000001 専門

0.000001 * 0.016282 * 100000 = 0.000927
0.000001 * 0.001468 * 100000 = 0.000083
ISDN/回線 0.000927 0.000083 専門

0.000051 * 0.000019 * 100000 = 0.000096
0.000001 * 0.000019 * 100000 = 0.000001
LAN/経由 0.000096 0.000001 専門

0.000001 * 0.004071 * 100000 = 0.000232
0.000001 * 0.000019 * 100000 = 0.000001
Mac/側 0.000232 0.000001 専門

0.000001 * 0.011759 * 100000 = 0.000670
0.000001 * 0.000220 * 100000 = 0.000013
PPP/接続 0.000670 0.000013 専門

0.000001 * 0.005880 * 100000 = 0.000335
0.000001 * 0.000367 * 100000 = 0.000021
USB/ポート 0.000335 0.000021 専門

0.000506 * 0.000823 * 0.016282 * 10000000 = 0.067719
0.000016 * 0.000003 * 0.001468 * 10000000 = 0.000001
アナログ/電話/回線 0.067719 0.000001 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000001 * 0.000294 * 100000 = 0.000017
キーノート/スピーチ 0.000001 0.000017 一般

0.000809 * 0.004071 * 100000 = 0.329297
0.000020 * 0.000294 * 100000 = 0.000585
ケーブル/モデム 0.329297 0.000585 専門

0.000253 * 0.001357 * 100000 = 0.034321
0.000042 * 0.000019 * 100000 = 0.000079
コピー/プロテクト 0.034321 0.000079 専門

0.000303 * 0.006332 * 100000 = 0.191890
0.000011 * 0.001615 * 100000 = 0.001745
コントロール/パネル 0.191890 0.001745 専門

0.000152 * 0.000019 * 100000 = 0.000288
0.000003 * 0.000294 * 100000 = 0.000084
シリアル/直結 0.000288 0.000084 専門

0.000001 * 0.025328 * 100000 = 0.001443
0.000022 * 0.000019 * 100000 = 0.000041
スタンド/型 0.001443 0.000041 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.000019 * 100000 = 0.000001

ダイヤルアップルーター/自体 0.000001 0.000001 専門

0.000303 * 0.004975 * 100000 = 0.150766

0.000023 * 0.000661 * 100000 = 0.001503

ダウン/ロード 0.150766 0.001503 専門

0.001264 * 0.013116 * 100000 = 1.657898

0.000047 * 0.000019 * 100000 = 0.000089

ネットワーク/環境 1.657898 0.000089 専門

0.000001 * 0.028042 * 100000 = 0.001597

0.000001 * 0.001064 * 100000 = 0.000060

ハンディ/端末 0.001597 0.000060 専門

0.000051 * 0.004975 * 100000 = 0.025222

0.000004 * 0.000019 * 100000 = 0.000008

マイクロソフト/社 0.025222 0.000008 専門

0.000051 * 0.005880 * 100000 = 0.029810

0.000001 * 0.000367 * 100000 = 0.000042

モデム/ポート 0.029810 0.000042 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000035 * 0.000019 * 100000 = 0.000066

ユーザー/自身 0.000001 0.000066 一般

0.000001 * 0.004071 * 100000 = 0.000232

0.000035 * 0.000019 * 100000 = 0.000066

ユーザー/側 0.000232 0.000066 専門

0.000455 * 0.000019 * 100000 = 0.000865

0.000001 * 0.000587 * 100000 = 0.000033

一般/ユーザー 0.000865 0.000033 専門

0.000152 * 0.000019 * 100000 = 0.000288

0.000001 * 0.000019 * 100000 = 0.000002
可能/性 0.000288 0.000002 専門

0.000202 * 0.000019 * 100000 = 0.000384
0.000060 * 0.000019 * 100000 = 0.000115
危険/性 0.000384 0.000115 専門

0.000253 * 0.000019 * 100000 = 0.000481
0.000001 * 0.000019 * 100000 = 0.000002
起動/時間 0.000481 0.000002 専門

0.000253 * 0.003166 * 100000 = 0.080074
0.000001 * 0.002239 * 100000 = 0.000255
起動/不能 0.080074 0.000255 専門

0.000101 * 0.039349 * 100000 = 0.396740
0.000023 * 0.000019 * 100000 = 0.000043
具体/的 0.396740 0.000043 専門

0.000051 * 0.000452 * 100000 = 0.002292
0.000001 * 0.000019 * 100000 = 0.000001
研究/機関 0.002292 0.000001 専門

0.000001 * 0.015830 * 100000 = 0.000902
0.000005 * 0.000019 * 100000 = 0.000010
見開き/表示 0.000902 0.000010 専門

0.000607 * 0.013116 * 100000 = 0.795701
0.000032 * 0.000019 * 100000 = 0.000062
更新/手続き 0.795701 0.000062 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000031 * 0.002606 * 100000 = 0.008001
講演/テーマ 0.000001 0.008001 一般

0.000455 * 0.016282 * 100000 = 0.741060
0.000007 * 0.000019 * 100000 = 0.000013

仕様/書 0.741060 0.000013 専門

$$0.000051 * 0.022162 * 100000 = 0.112356$$

$$0.000063 * 0.000697 * 100000 = 0.004399$$

手/入力 0.112356 0.004399 専門

$$0.000657 * 0.015378 * 100000 = 1.010892$$

$$0.000083 * 0.000019 * 100000 = 0.000158$$

周辺/技術 1.010892 0.000158 専門

$$0.000001 * 0.039349 * 100000 = 0.002241$$

$$0.000094 * 0.000019 * 100000 = 0.000178$$

商業/の 0.002241 0.000178 専門

$$0.000051 * 0.039349 * 100000 = 0.199491$$

$$0.000018 * 0.000019 * 100000 = 0.000033$$

消極/の 0.199491 0.000033 専門

$$0.000051 * 0.000019 * 100000 = 0.000096$$

$$0.000001 * 0.002018 * 100000 = 0.000115$$

新/機種 0.000096 0.000115 一般

$$0.000051 * 0.008593 * 100000 = 0.043565$$

$$0.000001 * 0.001284 * 100000 = 0.000073$$

新/仕様 0.043565 0.000073 専門

$$0.000101 * 0.000019 * 100000 = 0.000192$$

$$0.000038 * 0.000019 * 100000 = 0.000071$$

制限/時間 0.000192 0.000071 専門

$$0.000202 * 0.002714 * 100000 = 0.054883$$

$$0.000024 * 0.000019 * 100000 = 0.000045$$

設定/方法 0.054883 0.000045 専門

$$0.000001 * 0.017187 * 100000 = 0.000979$$

$$0.000007 * 0.000019 * 100000 = 0.000013$$

先駆/者 0.000979 0.000013 専門

0.000455 * 0.000118 * 0.028042 * 1000000 = 0.014988
0.000097 * 0.000054 * 0.001064 * 1000000 = 0.000056
専用/読書/端末 0.014988 0.000056 専門

0.000556 * 0.002714 * 100000 = 0.150889
0.000030 * 0.000019 * 100000 = 0.000056
接続/方法 0.150889 0.000056 専門

0.000051 * 0.000019 * 100000 = 0.000096
0.000043 * 0.000019 * 100000 = 0.000081
対応/策 0.000096 0.000081 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000009 * 0.000019 * 100000 = 0.000017
対処/法 0.000001 0.000017 一般

0.000001 * 0.003618 * 100000 = 0.000206
0.000026 * 0.000019 * 100000 = 0.000049
貸し出し/作業 0.000206 0.000049 専門

0.000202 * 0.000019 * 100000 = 0.000384
0.000001 * 0.000019 * 100000 = 0.000001
大/企業 0.000384 0.000001 専門

0.000253 * 0.000013 * 0.000019 * 1000000 = 0.000001
0.000108 * 0.000003 * 0.000073 * 1000000 = 0.000000
通常/IP/アド 0.000001 0.000000 専門

0.001466 * 0.013116 * 100000 = 1.923131
0.000001 * 0.000019 * 100000 = 0.000001
通信/環境 1.923131 0.000001 専門

0.000152 * 0.001809 * 100000 = 0.027411
0.000111 * 0.000019 * 100000 = 0.000212
低/コスト 0.027411 0.000212 専門

$$0.001820 * 0.002714 * 100000 = 0.493946$$

$$0.000105 * 0.000954 * 100000 = 0.009981$$

電子/ブック 0.493946 0.009981 専門

$$0.001466 * 0.000013 * 0.000019 * 10000000 = 0.000004$$

$$0.000001 * 0.000027 * 0.001908 * 10000000 = 0.000000$$

電子/ブック/リーダー 0.000004 0.000000 専門

$$0.001466 * 0.000013 * 0.003618 * 10000000 = 0.000709$$

$$0.000001 * 0.000027 * 0.000019 * 10000000 = 0.000000$$

電子/ブック/用 0.000709 0.000000 専門

$$0.001820 * 0.007689 * 100000 = 1.399393$$

$$0.000105 * 0.000624 * 100000 = 0.006528$$

電子/辞書 1.399393 0.006528 専門

$$0.000303 * 0.000019 * 100000 = 0.000576$$

$$0.000001 * 0.000019 * 100000 = 0.000001$$

同/アップデーター 0.000576 0.000001 専門

$$0.000001 * 0.028042 * 100000 = 0.001597$$

$$0.000049 * 0.001064 * 100000 = 0.005263$$

読書/端末 0.001597 0.005263 一般

$$0.000961 * 0.003172 * 0.000019 * 10000000 = 0.000579$$

$$0.000001 * 0.000003 * 0.000019 * 10000000 = 0.000000$$

日本/語/OS 0.000579 0.000000 専門

$$0.000051 * 0.000019 * 100000 = 0.000096$$

$$0.000001 * 0.000019 * 100000 = 0.000001$$

販売/価格 0.000096 0.000001 専門

$$0.001820 * 0.000452 * 100000 = 0.082264$$

$$0.000030 * 0.000183 * 100000 = 0.000541$$

表示/液晶 0.082264 0.000541 専門

$$0.001820 * 0.000452 * 100000 = 0.082264$$

$$0.000030 * 0.000037 * 100000 = 0.000109$$

表示/解像度 0.082264 0.000109 専門

$$0.000050 * 0.005051 * 0.000019 * 10000000 = 0.000048$$

$$0.000001 * 0.000169 * 0.000019 * 10000000 = 0.000000$$

連続/表示/時間 0.000048 0.000000 専門

文書 5

本文

昨年来、常に刺激的な話題を提供し続けている、インターネット音楽配信の分野だが、先ごろ、またしても興味深い事件が起こった。

MP3による音楽配信の中核的サイトである「MP3.com」を、米レコード協会 (RIAA) がニューヨーク連邦地裁に著作権侵害で提訴したのである。

問題にされたのは、「MP3.com」が2000年1月に同サイトのパーソナライズ・サービス「My.MP3.com」の一環としてスタートしたばかりの新サービス「Beam-it」と「Instant Listening Service」だ。

「Beam-it」は、ローカルのさまざまな音源を統括管理するいわゆるジュークボックス・ソフトをオンライン・アプリケーションとして拡張したようなサービス。手持ちのCDをパソコンのCD-ROMドライブに入れると、「Beam-it」のクライアント・アプリケーションは、「Beam-it」のサーバに、そのCDのデータが存在するかを見に行く。存在すれば、そのCDの楽曲リストが「My.MP3.com」の自分用のページに表示されるようになる。そしてユニークなのは、これ以降、CDをドライブから取り出しても、そのCDの音楽を「My.MP3.com」からのストリーミングで聴くことができるようになることだ。これによってユーザは、どこにいても、「My.MP3.com」にアクセスする環境があれば、自分の持っているCDを聴くことができることになる。

もうひとつの「Instant Listening Service」は、「MP3.com」と提携しているオンラインCDショップでCDを購入すると、そのCDが自宅に届く前に、CDの音楽をストリーミングで聴けるようになるというサービスだ。

音楽配信の事情を少しでも知っている人ならすぐわかると思うが、これはいずれも非常に面白いと同時に、なかなかキワどいサービスでもある。CDはリスナーが購入しているわけだから、著作権者の金銭面における権利を直接的に侵害はしていない。しかし、著作権者の許諾なく、音楽作品を配信していることには変わりない、と見れば見れないこともない。法解釈によってはクロの可能性もあるが、誰も実害を被らず、ユーザにはベネフィットがあるのだからいいのではないか、という主張もあるだろう。逆に、レコード会社からすると、オフィシャルの音楽配信のためのデータベースを準備している段階で、勝手に自分とこの音源を配信されたのではたまらんといいがあるだろう。

この事件は、デジタルの音楽ソフトはもはや、CDのポリカーボネイトに閉じ込めておくものではない、自宅のCD棚に並べておくよりも、もっと便利で自由なリスニング・スタイルが可能だ、ということのゲリラ的・確信犯的な実践と言えるのではないだ

[1] line:1 byte:2319/4415 52%

ろうか。あるいは、未だ本格始動しないメジャーなレコード会社の音楽配信ビジネスに対するリスナーのいらだちを代弁するものとも言えるだろう。

本格的な音楽配信の時代を前に、音楽ファンのインターネット・サービス・プロバイダ（ISP）選びにも新しい基準が求められるようになる。

値段がリーズナブルであるのはもちろんだが、従量制より定額制が適している。たとえばストリーミングでインターネット・ライブを楽しんでいるときに、ISPの課金が気になったのでは台無しだ。

信頼できる回線の品質も重要だ。同じ速度のモデム／TAを使っている、回線品質が悪いと、ストリーミングならば音が途切れることが多くなるだろうし、ダウンロードには余計に時間がかかってしまう。いい音質で快適に音楽配信を楽しむには、単にできる限り安くメールが読めればオッケーという場合とは違う選択眼が必要だ。

また、契約するISPが独自でどのようなコンテンツ／サービスを用意しているか、ということも忘れずチェックしたい。近頃は、差別化要素としてオリジナルのコンテンツを拡充するISPも多いが、音楽ファンならば、当然、充実した音楽コンテンツ／サービスを提供しているISPの方がメリットが大きいのは言うまでもないだろう。

PURITY ONLINE (PUON) は、大阪有線株式会社が運営するISPである。

このISPの特徴として真っ先に目をひくのは、\2,500／月（\25,000／年）というリーズナブルな定額制でありながら、回線品質が高いということ。バックボーンネットワークとして世界的に高品質が認められている「UUNET」が使われているのだ。

さらに、「元祖・音楽配信」とも言えるビジネスを展開してきた企業ならではの個性がPUONにはある。音楽コンテンツが充実しているのだ。

PUONは独自の音楽ポータル・サイト「ミュージック・プレジャー・オンライン」を運営している。「ミュージック・プレジャー・オンライン」では大阪有線の有線放送プログラム「Usen440」の中の10チャンネルがストリーミング配信されている他、音楽ファンのための掲示板や、音楽を聴きたい人演りたい人のための最新情報などが提供されている。

また、PUONの会員には、カラオケ・ダウンロード・サービス「PCユーカラ」が毎月1500円分まで無料という特典もある。

ISPの選択肢はそれこそ星の数ほどある。だからこそ、冷静に自分に合ったISPを選ぶ必要がある。あなたが音楽ファンならば、たとえばPUONには検討の余地があるのではないだろうか。

判別結果

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.001872 * 100000 = 0.000106

10/チャンネル 0.000001 0.000106 一般

0.000001 * 0.000118 * 0.000019 * 10000000 = 0.000000

0.000001 * 0.000076 * 0.000019 * 10000000 = 0.000000

1500/円/分 0.000000 0.000000 専門

0.000001 * 0.002261 * 100000 = 0.000129

0.000001 * 0.000734 * 100000 = 0.000042

CD-ROM/ドライブ 0.000129 0.000042 専門

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.000550 * 100000 = 0.000031

CD/棚 0.000001 0.000031 一般

0.000001 * 0.000019 * 100000 = 0.000001

0.000001 * 0.000037 * 100000 = 0.000002

PC/ユーカラ 0.000001 0.000002 一般

0.000001 * 0.000118 * 0.000452 * 10000000 = 0.000000

0.000009 * 0.000003 * 0.000404 * 10000000 = 0.000000

インターネット/音楽/配信 0.000000 0.000000 専門

0.000001 * 0.039349 * 100000 = 0.002241

0.000049 * 0.000019 * 100000 = 0.000094

ゲリラ/的 0.002241 0.000094 専門

0.000051 * 0.000452 * 100000 = 0.002292

0.000001 * 0.000404 * 100000 = 0.000023

ストリーミング/配信 0.002292 0.000023 専門

0.000303 * 0.004975 * 100000 = 0.150766

0.000023 * 0.000661 * 100000 = 0.001503

ダウン/ロード 0.150766 0.001503 専門

0.000001 * 0.000013 * 0.000019 * 10000000 = 0.000000
0.000193 * 0.000328 * 0.003303 * 10000000 = 0.002093
ニューヨーク/連邦/地裁 0.000000 0.002093 一般

0.000101 * 0.000019 * 100000 = 0.000192
0.000024 * 0.000514 * 100000 = 0.001227
ブレ/ジャー 0.000192 0.001227 一般

0.000758 * 0.000013 * 0.000452 * 10000000 = 0.000046
0.000012 * 0.000003 * 0.002606 * 10000000 = 0.000001
オンライン/CD/ショップ 0.000046 0.000001 専門

0.000152 * 0.003166 * 100000 = 0.047972
0.000038 * 0.000019 * 100000 = 0.000071
レコード/会社 0.047972 0.000071 専門

0.000101 * 0.000019 * 100000 = 0.000192
0.000001 * 0.000019 * 100000 = 0.000001
音楽/コンテンツ 0.000192 0.000001 専門

0.000101 * 0.001809 * 100000 = 0.018239
0.000001 * 0.000019 * 100000 = 0.000001
音楽/ソフト 0.018239 0.000001 専門

0.000101 * 0.000019 * 100000 = 0.000192
0.000001 * 0.000019 * 100000 = 0.000001
音楽/ファン 0.000192 0.000001 専門

0.000101 * 0.000019 * 100000 = 0.000192
0.000001 * 0.000019 * 100000 = 0.000001
音楽/ポータル 0.000192 0.000001 専門

0.000101 * 0.000019 * 100000 = 0.000192
0.000001 * 0.000019 * 100000 = 0.000001
音楽/作品 0.000192 0.000001 専門

0.000101 * 0.000452 * 100000 = 0.004557

0.000001 * 0.000404 * 100000 = 0.000023

音楽/配信 0.004557 0.000023 専門

0.000001 * 0.000013 * 0.000019 * 10000000 = 0.000000

0.000001 * 0.000016 * 0.003633 * 10000000 = 0.000001

音楽/配信/ビジネス 0.000000 0.000001 一般

0.000152 * 0.000019 * 100000 = 0.000288

0.000001 * 0.000019 * 100000 = 0.000002

可能/性 0.000288 0.000002 専門

0.001213 * 0.002714 * 100000 = 0.329298

0.000014 * 0.000440 * 100000 = 0.000600

回線/品質 0.329298 0.000600 専門

0.000001 * 0.000013 * 0.039349 * 10000000 = 0.000005

0.000005 * 0.000126 * 0.000019 * 10000000 = 0.000000

確信/犯/的 0.000005 0.000000 専門

0.000001 * 0.021710 * 100000 = 0.001237

0.000008 * 0.000019 * 100000 = 0.000015

楽曲/リスト 0.001237 0.000015 専門

0.000001 * 0.004071 * 100000 = 0.000232

0.000044 * 0.000019 * 100000 = 0.000084

金銭/面 0.000232 0.000084 専門

0.000202 * 0.002714 * 100000 = 0.054883

0.000001 * 0.000440 * 100000 = 0.000025

高/品質 0.054883 0.000025 専門

0.000001 * 0.009163 * 0.019448 * 10000000 = 0.001534

0.000073 * 0.000003 * 0.002679 * 10000000 = 0.000005

差別/化/要素 0.001534 0.000005 専門

0.000051 * 0.015378 * 100000 = 0.077963

0.000090 * 0.000019 * 100000 = 0.000171
最新/情報 0.077963 0.000171 専門

0.000001 * 0.039349 * 100000 = 0.002241
0.000018 * 0.000019 * 100000 = 0.000033
刺激/的 0.002241 0.000033 専門

0.000001 * 0.003618 * 100000 = 0.000206
0.000038 * 0.000019 * 100000 = 0.000071
自分/用 0.000206 0.000071 専門

0.000051 * 0.028494 * 100000 = 0.144458
0.000001 * 0.000019 * 100000 = 0.000001
新/サービス 0.144458 0.000001 専門

0.000101 * 0.039349 * 100000 = 0.396740
0.000001 * 0.000019 * 100000 = 0.000001
世界/的 0.396740 0.000001 専門

0.001416 * 0.000019 * 100000 = 0.002690
0.000045 * 0.000550 * 100000 = 0.002502
選択/眼 0.002690 0.002502 専門

0.000101 * 0.010925 * 0.000905 * 10000000 = 0.009998
0.000045 * 0.000003 * 0.000183 * 10000000 = 0.000000
中核/的/サイト 0.009998 0.000000 専門

0.000101 * 0.000470 * 0.017187 * 10000000 = 0.008170
0.000080 * 0.000003 * 0.000019 * 10000000 = 0.000000
著作/権/者 0.008170 0.000000 専門

0.000101 * 0.000470 * 0.000019 * 10000000 = 0.000009
0.000080 * 0.000003 * 0.000771 * 10000000 = 0.000002
著作/権/侵害 0.000009 0.000002 専門

0.000001 * 0.000019 * 100000 = 0.000001
0.000015 * 0.000019 * 100000 = 0.000028

定額/制 0.000001 0.000028 一般

$$0.000001 * 0.036183 * 100000 = 0.002061$$

$$0.000018 * 0.000019 * 100000 = 0.000033$$

統括/管理 0.002061 0.000033 専門

$$0.000303 * 0.000905 * 100000 = 0.027426$$

$$0.000001 * 0.000183 * 100000 = 0.000010$$

同/サイト 0.027426 0.000010 専門

$$0.000001 * 0.001644 * 0.002261 * 10000000 = 0.000032$$

$$0.000001 * 0.000060 * 0.000019 * 10000000 = 0.000000$$

米/レコード/協会 0.000032 0.000000 専門

$$0.000253 * 0.000452 * 100000 = 0.011432$$

$$0.000074 * 0.001798 * 100000 = 0.013290$$

法/解釈 0.011432 0.013290 一般

$$0.000001 * 0.000019 * 100000 = 0.000001$$

$$0.000001 * 0.000183 * 100000 = 0.000010$$

本格/始動 0.000001 0.000010 一般

$$0.000001 * 0.039349 * 100000 = 0.002241$$

$$0.000001 * 0.000019 * 100000 = 0.000001$$

本格/的 0.002241 0.000001 専門

$$0.000152 * 0.004071 * 100000 = 0.061685$$

$$0.000009 * 0.000019 * 100000 = 0.000017$$

有線/放送 0.061685 0.000017 専門

プログラムリスト

複合語抽出プログラム

```
#include<stdio.h>
#include<stdlib.h>
#include<string.h>
#include<ctype.h>
main(int argc,char *argv[])
{
    FILE *f1;
    char buf[1000];
    char buf1[1000];
    char buf2[1000],buf5[1000];
    char buf3[1000],buf4[1000];
    char *a,*b,*c,*d,*e,*f,*g,*h;
    char aa[]="名詞\0";
    char bb[]="数\0";
    char cc[]="助数詞\0";
    char dd[]="副詞可能\0";
    /* char ee[]="接尾\0";*/
    char ff[]="代名詞\0";
    char gg[]="人名\0";
    char hh[]="非自立\0";

    int r,n=0;int mm=0;
    int v,w,x,y,z,o,p;
    int i=0;int j=0;
    int num,num2=0;
    for(mm=1; mm < argc; mm++) {
```

```

/*if(argc !=2) {printf("引数の数が違う\n"); exit(1); }*/
if((f1 = fopen(argv[mm],"r"))==NULL) exit(1);

while(fgets(buf,1000,f1)!=NULL)
{

    a=strstr(buf,aa);
    b=strstr(buf,bb);
    c=strstr(buf,cc);
    d=strstr(buf,dd);
    /*e=strstr(buf,ee);*/
    f=strstr(buf,ff);
    g=strstr(buf,gg);
    h=strstr(buf,hh);

    if(a!=NULL)
    {
        j++;
        if(b==NULL&&c==NULL&&d==NULL&&f==NULL&&g==NULL&&h==NULL)
        {
            if(j==5){strcpy(buf5,buf);}
            if(j==4){strcpy(buf4,buf);}
            if(j==3){strcpy(buf3,buf);}
            if(j==2){strcpy(buf2,buf);}
            if(j==1){strcpy(buf1,buf);}
        }
        else{j=0;}
    }
    else{num=j;j=0;}

    if(num==5)
    {

```

```

for(i=0;i<=strlen(buf1);i++)
{
    if(isspace(buf1[i])==0)
        {printf("%c",buf1[i]);}
    else{i=0;break;}
}
for(i=0;i<=strlen(buf2);i++)
{
    if(isspace(buf2[i])==0)
        {printf("%c",buf2[i]);}
    else{i=0;break;}
}
for(i=0;i<=strlen(buf3);i++)
{
    if(isspace(buf3[i])==0)
        {printf("%c",buf3[i]);}
    else{i=0;break;}
}
for(i=0;i<=strlen(buf4);i++)
{
    if(isspace(buf4[i])==0)
        {printf("%c",buf4[i]);}
    else{i=0;break;}
}
for(i=0;i<=strlen(buf5);i++)
{
    if(isspace(buf5[i])==0)
        {printf("%c",buf5[i]);}
    else{printf("\n");i=0;num=0;break;}
}
}

```

```

if(num==4)
{
    for(i=0;i<=strlen(buf1);i++)

```

```

    {
        if(isspace(buf1[i])==0)
            {printf("%c",buf1[i]);}
        else{i=0;break;}
    }
for(i=0;i<=strlen(buf2);i++)
    {
        if(isspace(buf2[i])==0)
            {printf("%c",buf2[i]);}
        else{i=0;break;}
    }
for(i=0;i<=strlen(buf3);i++)
    {
        if(isspace(buf3[i])==0)
            {printf("%c",buf3[i]);}
        else{i=0;break;}
    }
for(i=0;i<=strlen(buf4);i++)
    {
        if(isspace(buf4[i])==0)
            {printf("%c",buf4[i]);}
        else{printf("\n");i=0;num=0;break;}
    }
}

```

```

if(num==3)
{
    for(i=0;i<=strlen(buf1);i++)
        {
            if(isspace(buf1[i])==0)
                {printf("%c",buf1[i]);}
            else{i=0;break;}
        }
    for(i=0;i<=strlen(buf2);i++)

```

```

        {
            if(isspace(buf2[i])==0)
                {printf("%c",buf2[i]);}
            else{i=0;break;}
        }
    for(i=0;i<=strlen(buf3);i++)
        {
            if(isspace(buf3[i])==0)
                {printf("%c",buf3[i]);}
            else{printf("\n");i=0;num=0;break;}
        }
    }

```

```

if(num==2)
{
    for(i=0;i<=strlen(buf1);i++)
        {
            if(isspace(buf1[i])==0)
                {printf("%c",buf1[i]);}
            else{i=0;break;}
        }
    for(i=0;i<=strlen(buf2);i++)
        {
            if(isspace(buf2[i])==0)
                {printf("%c",buf2[i]);}
            else{printf("\n");i=0;num=0;break;}
        }
    }
}

```

```

fclose(f1);
    }
}

```

専門用語判別プログラム

```
#include<stdio.h>
#include<stdlib.h>
#include<string.h>
#include<ctype.h>

char buf[1000],buf2[1000],buf6[1000];
char buf3[1000],buf4[1000],buf5[1000];

FILE *f1,*f2;
int bs =0,check=0,i=0,j=0,count,k=0;

float kakuritu();

float keta1,keta2,keta3,keta4,keta5;
float ritu,ritu1,ritu2,ritu3,ritu4,ritu5,ritu6;

main(int argc,char *argv[])
{

float kekka1,kekka2;
char info1[100],info2[100],info3[100],info4[100],info5[100];
char aa[]="EOS\0";char *a;

int r,n=0;int mm;
float x,y;

for(mm=1; mm < argc; mm++)
{

if((f1 = fopen(argv[mm],"r"))==NULL) exit(1);
while(fgets(buf,1000,f1)!=NULL)
{

if(i==0){strcpy(buf2,buf);}


```

```

if(i==1){strcpy(buf3,buf);}
if(i==2){strcpy(buf4,buf);}
if(i==3){strcpy(buf5,buf);}
if(i==4){strcpy(buf6,buf);}

```

```

a=strstr(buf,aa);
if(a!=NULL)
{

```

```

if(i==2)
{

```

```

/*-----情報-----*/
f2=fopen("info.nigo.kazu2","r");
for(bs=0;bs<=strlen(buf2);bs++)
{if(isspace(buf2[bs])!=0){break;}}
check=0;
ritu1 = kakuritu();
ritu1 = ritu1 * 0.569638; /*printf("%f\n",ritu1);*/
if(count!=1){ritu1=0.000001 * 0.569638;count=0;}

```

```

f2=fopen("info.end.kazu1","r");
for(bs=0;bs<=strlen(buf3);bs++)
{if(isspace(buf3[bs])!=0){break;}}
check=1;
ritu2=kakuritu(); /*printf("%f\n",ritu2);*/
if(count!=1){ritu2=0.000019;count=0;}

```

```

/*----- 確率 -----*/
kekka1 = ritu1 * ritu2 * 100000;
printf("\n %f * %f * 100000 = %f\n",ritu1,ritu2,kekka1);
/*printf("専門確率 = %f\n\n",ritu3);*/

```

```

/*----- 一般 -----*/
f2=fopen("ippango.nigo1.kazu2","r");
for(bs=0;bs<=strlen(buf2);bs++)
    {if(isspace(buf2[bs])!=0){
        strncpy(info1,buf2,bs);
        info1[bs]='/';info1[bs+1]='\0';break;}}
check=0;
ritu1 = kakuritu();
ritu1 = ritu1 * 0.5685921 ;/*printf("%f\n",ritu1);*/
if(count!=1){ritu1=0.5685921*0.000001;count=0;}

f2=fopen("ippango.end.kazu1","r");
for(bs=0;bs<=strlen(buf3);bs++)
    {if(isspace(buf3[bs])!=0){
        strncpy(info2,buf3,bs);
        info2[bs]='\0';break;}}
check=1;
ritu2=kakuritu();/*printf("%f\n",ritu2);*/
if(count!=1){ritu2=0.000019;count=0;}

/*----- 確率 -----*/
kekka2=ritu1*ritu2*100000;
printf(" %f * %f * 100000 = %f\n",ritu1,ritu2,kekka2);

strcat(info1,info2);printf("%s ",info1);

printf("%f %f",kekka1,kekka2);
if(kekka1>kekka2){printf(" 専門\n");}
else{printf(" 一般\n");}
}

```

```

if(i==3)
{
/*-----情報-----*/
f2=fopen("info.sentou.kazu2","r");
for(bs=0;bs<=strlen(buf2);bs++)
{if(isspace(buf2[bs])!=0){break;}}
check=0;
ritu1 = kakuritu();
ritu1 = ritu1 * 0.430336247;
if(count!=1){ritu1=0.000002 * 0.430336247;count=0;}

f2=fopen("info.aida.end.kazu2","r");
for(bs=0;bs<=strlen(buf3);bs++)
{if(isspace(buf3[bs])!=0){break;}}
check=1;
ritu2=kakuritu();
ritu2=ritu2 * 0.70362974;
if(count!=1){ritu2=0.000019 * 0.70362974;count=0;}

f2=fopen("info.end.kazu1","r");
for(bs=0;bs<=strlen(buf4);bs++)
{if(isspace(buf4[bs])!=0){break;}}
check=2;
ritu3=kakuritu();
if(count!=1){ritu3=0.000019;count=0;}

/*----- 確率 -----*/
kekka1 = ritu1 * ritu2 * ritu3 *10000000;
printf("\n %f * %f * %f * 10000000 =
%f\n",ritu1,ritu2,ritu3,kekka1);
/*printf("専門確率 = %f\n\n",ritu4);*/

/*----- 一般 -----*/
f2=fopen("ippango.sentou.kazu2","r");
for(bs=0;bs<=strlen(buf2);bs++)

```

```

        {if(isspace(buf2[bs])!=0){
            strncpy(info1,buf2,bs);
            info1[bs]='/' ;info1[bs+1]='\0';break;}}
    check=0;
    ritu1 = kakuritu();
    ritu1 = ritu1 * 0.4314079;
    if(count!=1){ritu1=0.4314079 * 0.000002 ;count=0;}

    f2=fopen("ippango.aida.end.kazu2","r");
    for(bs=0;bs<=strlen(buf3);bs++)
        {if(isspace(buf3[bs])!=0){
            strncpy(info2,buf3,bs);
            info2[bs]='/' ;info2[bs+1]='\0';break;}}
    check=1;
    ritu2=kakuritu();
    ritu2 = ritu2 * 0.68023316 ;
    if(count!=1){ritu2=0.000004 * 0.68023316;count=0;}

    f2=fopen("ippango.end.kazu1","r");
    for(bs=0;bs<=strlen(buf4);bs++)
        {if(isspace(buf4[bs])!=0){
            strncpy(info3,buf4,bs);
            info3[bs]='\0';break;}}
    check=2;
    ritu3=kakuritu();
    if(count!=1){ritu3=0.000019;count=0;}

    /*----- 確率 -----*/
    kekka2=ritu1*ritu2*ritu3*10000000;
    printf(" %f * %f * %f * 10000000 = %f\n",ritu1,ritu2,ritu3,kekka2);

    strcat(info1,info2);strcat(info1,info3);
    printf("%s ",info1);

```

```

printf("%f %f",kekka1,kekka2);
if(kekka1>kekka2){printf(" 専門\n");}
else{printf(" 一般\n");}
}

```

```

if(i==4)
{
/*-----情報-----*/
f2=fopen("info.sentou.kazu2","r");
for(bs=0;bs<=strlen(buf2);bs++)
{if(isspace(buf2[bs])!=0){break;}}
check=0;
ritu1 = kakuritu();
ritu1 = ritu1 * 0.430336247;
if(count!=1){ritu1=0.000002 * 0.430336247;count=0;}

f2=fopen("info.aida.kazu2","r");
for(bs=0;bs<=strlen(buf3);bs++)
{if(isspace(buf3[bs])!=0){break;}}
check=1;
ritu2=kakuritu();
ritu2=ritu2 * 0.29637026;
if(count!=1){ritu2=0.000009 * 0.29637026;count=0;}

f2=fopen("info.aida.end.kazu2","r");
for(bs=0;bs<=strlen(buf4);bs++)
{if(isspace(buf4[bs])!=0){break;}}
check=2;
ritu3=kakuritu();
ritu3=ritu3 * 0.70362974;

```

```

if(count!=1){ritu3=0.000004 * 0.70362974;count=0;}

f2=fopen("info.end.kazu1","r");
for(bs=0;bs<=strlen(buf5);bs++)
    {if(isspace(buf5[bs])!=0){break;}}
check=3;
ritu4=kakuritu();
if(count!=1){ritu4=0.000019;count=0;}

/*----- 確率 -----*/
kekka1 = ritu1 * ritu2 * ritu3 * ritu4 * 1000000000;
printf("\n %f * %f * %f * %f * 1000000000 = %f\n"
        ,ritu1,ritu2,ritu3,ritu4,kekka1);
/*printf("専門確率 = %f\n\n",kekka1);*/

/*----- 一般 -----*/
f2=fopen("ippango.sentou.kazu2","r");
for(bs=0;bs<=strlen(buf2);bs++)
    {if(isspace(buf2[bs])!=0){
        strncpy(info1,buf2,bs);
        info1[bs]='/';info1[bs+1]='\0';break;}}
check=0;
ritu1 = kakuritu();
ritu1 = ritu1 * 0.4314079 ;
if(count!=1){ritu1=0.4314079 * 0.000002 ;count=0;}

f2=fopen("ippango.aida.kazu2","r");
for(bs=0;bs<=strlen(buf3);bs++)
    {if(isspace(buf3[bs])!=0){
        strncpy(info2,buf3,bs);
        info2[bs]='/';info2[bs+1]='\0';break;}}
check=1;
ritu2=kakuritu();
ritu2 = ritu2 * 0.31976684;
if(count!=1){ritu2=0.000004 * 0.31976684;count=0;}

```

```

f2=fopen("ippango.aida.end.kazu2","r");
for(bs=0;bs<=strlen(buf4);bs++)
    {if(isspace(buf4[bs])!=0){
        strncpy(info3,buf4,bs);
        info3[bs]='/' ;info3[bs+1]='\0' ;break;}}
check=2;
ritu3=kakuritu();
ritu3 = ritu3 * 0.68023316 ;
if(count!=1){ritu3=0.000004 * 0.68023316;count=0;}

```

```

f2=fopen("ippango.end.kazu1","r");
for(bs=0;bs<=strlen(buf5);bs++)
    {if(isspace(buf5[bs])!=0){
        strncpy(info4,buf5,bs);
        info4[bs]='\0' ;break;}}
check=3;
ritu4=kakuritu();
if(count!=1){ritu4=0.000019;count=0;}

```

```

/*----- 確率 -----*/
kekka2=ritu1*ritu2*ritu3*ritu4*1000000000;
printf(" %f * %f * %f * %f * 1000000000 = %f\n"
        ,ritu1,ritu2,ritu3,ritu4,kekka2);

```

```

strcat(info1,info2);strcat(info1,info3);strcat(info1,info4);
printf("%s ",info1);

```

```

printf("%f %f",kekka1,kekka2);
if(kekka1>kekka2){printf(" 専門\n");}
else{(" 一般\n");}
}

```

```

if(i==5)
{
/*-----情報-----*/
f2=fopen("info.sentou.kazu2","r");
for(bs=0;bs<=strlen(buf2);bs++)
    {if(isspace(buf2[bs])!=0){break;}}
check=0;
ritu1 = kakuritu();
ritu1 = ritu1 * 0.430336247;
if(count!=1){ritu1=0.000002 * 0.430336247;count=0;}

f2=fopen("info.aida.kazu2","r");
for(bs=0;bs<=strlen(buf3);bs++)
    {if(isspace(buf3[bs])!=0){break;}}
check=1;
ritu2=kakuritu();
ritu2=ritu2 * 0.29637026;
if(count!=1){ritu2=0.000009 * 0.29637026;count=0;}

f2=fopen("info.aida.kazu2","r");
for(bs=0;bs<=strlen(buf4);bs++)
    {if(isspace(buf4[bs])!=0){break;}}
check=2;
ritu3=kakuritu();
ritu3=ritu3 * 0.29637026;
if(count!=1){ritu3=0.000009 * 0.29637026;count=0;}

f2=fopen("info.aida.end.kazu2","r");
for(bs=0;bs<=strlen(buf5);bs++)
    {if(isspace(buf5[bs])!=0){break;}}
check=3;
ritu4=kakuritu();
ritu4=ritu4 * 0.70362974;
if(count!=1){ritu4=0.000004 * 0.70362974;count=0;}

```

```

f2=fopen("info.end.kazu1","r");
for(bs=0;bs<=strlen(buf6);bs++)
    {if(isspace(buf6[bs])!=0){break;}}
check=4;
ritu5=kakuritu();
if(count!=1){ritu5=0.000019;count=0;}

/*----- 確率 -----*/
kekka1 = ritu1 * ritu2 * ritu3 * ritu4 * ritu5 * 1000000000;
printf("\n %f * %f * %f * %f * %f * 1000000000 = %f\n"
    ,ritu1,ritu2,ritu3,ritu4,kekka1);
/*printf("専門確率 = %f\n\n",kekka1);*/

```

```

/*----- 一般 -----*/
f2=fopen("ippango.sentou.kazu2","r");
for(bs=0;bs<=strlen(buf2);bs++)
    {if(isspace(buf2[bs])!=0){
        strncpy(info1,buf2,bs);
        info1[bs]='/';info1[bs+1]='\0';break;}}
check=0;
ritu1 = kakuritu();
ritu1 = ritu1 * 0.4314079;
if(count!=1){ritu1=0.4314079 * 0.000002;count=0;}

```

```

f2=fopen("ippango.aida.kazu2","r");
for(bs=0;bs<=strlen(buf3);bs++)
    {if(isspace(buf3[bs])!=0){
        strncpy(info2,buf3,bs);
        info2[bs]='/';info2[bs+1]='\0';break;}}
check=1;
ritu2=kakuritu();
ritu2 = ritu2 * 0.31976684;
if(count!=1){ritu2=0.000004 * 0.31976684;count=0;}

```

```

f2=fopen("ippango.aida.kazu2","r");
for(bs=0;bs<=strlen(buf4);bs++)

```

```

        {if(isspace(buf4[bs])!=0){
            strncpy(info3,buf4,bs);
            info3[bs]='/';info3[bs+1]='\0';break;}}
    check=2;
    ritu3=kakuritu();
    ritu3 = ritu3 * 0.31976684;
    if(count!=1){ritu3=0.000004 * 0.31976684;count=0;}

    f2=fopen("ippango.aida.end.kazu2","r");
    for(bs=0;bs<=strlen(buf5);bs++)
        {if(isspace(buf5[bs])!=0){
            strncpy(info4,buf5,bs);
            info4[bs]='/';info4[bs+1]='\0';break;}}
    check=3;
    ritu4=kakuritu();
    ritu4 = ritu4 * 0.68023316;
    if(count!=1){ritu4=0.000004 * 0.68023316;count=0;}

    f2=fopen("ippango.end.kazu1","r");
    for(bs=0;bs<=strlen(buf6);bs++)
        {if(isspace(buf6[bs])!=0){
            strncpy(info5,buf6,bs);
            info5[bs]='\0';break;}}
    check=4;
    ritu5=kakuritu();
    if(count!=1){ritu5=0.000019;count=0;}

    /*----- 確率 -----*/
    kekka2=ritu1*ritu2*ritu3*ritu4*ritu5*1000000000;
    printf(" %f * %f * %f * %f * %f * 1000000000 = %f\n"
        ,ritu1,ritu2,ritu3,ritu4,kekka1);

    strcat(info1,info2);strcat(info1,info3);strcat(info1,info4);

```

```

        strcat(info1,info5);printf("%s ",info1);

        printf("%f %f",kekka1,kekka2);
        if(kekka1>kekka2){printf(" 専門\n");}
        else{printf(" 一般\n");}
            }

        }
        else{i++;}
    }
    fclose(f1);
}
}

```

```

/*----- 確率計算 -----*/

```

```

float kakuritu()
{

    int cmp1;

    while(fgets(buf,1000,f2)!=NULL)
    {
        if(check==0){cmp1=strncmp(buf2,buf,bs);}
        if(check==1){cmp1=strncmp(buf3,buf,bs);}
        if(check==2){cmp1=strncmp(buf4,buf,bs);}
        if(check==3){cmp1=strncmp(buf5,buf,bs);}
        if(check==4){cmp1=strncmp(buf6,buf,bs);}

        if((cmp1==0)&&(isspace(buf[bs])!=0))
        {

            for(j=0;j<=strlen(buf);j++)
            {

```

```

if (isspace(buf[j])!=0){k++;}
if (k==4)
{
    if (j==strlen(buf)-6)
        {keta5=buf[j];keta5=(keta5-48)*0.01;}
    if (j==strlen(buf)-5)
        {keta4=buf[j];keta4=(keta4-48)*0.001;}
    if (j==strlen(buf)-4)
        {keta3=buf[j];keta3=(keta3-48)*0.0001;}
    if (j==strlen(buf)-3)
        {keta2=buf[j];keta2=(keta2-48)*0.00001;}
    if (j==strlen(buf)-2)
        {keta1=buf[j];keta1=(keta1-48)*0.000001;}
}
}

```

```

ritu = keta5+keta4+keta3+keta2+keta1;

```

```

k=0;
count=1;i=0;
return(ritu);
}

```

```

}

```

```

fclose(f2);i=0;count=0;
}

```